PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C12N 15/12, 15/62, C07K 13/00
C12N 9/16, 9/18, 9/24
C12N 9/64, A61K 37/02

A1

(11) International Publication Number: WO 93/06216

(43) International Publication Date: 1 April 1993 (01.04.93)

(21) International Application Number: PCT/US92/08090

(22) International Filing Date: 22 September 1992 (22.09.92)

(30) Priority data:
766,963 26 September 1991 (26.09.91) US

(71) Applicant: OKLAHOMA MEDICAL RESEARCH FOUNDATION [US/US]; 825 N.E. Thirteenth Street, Oklahoma City, OK 73104 (US).

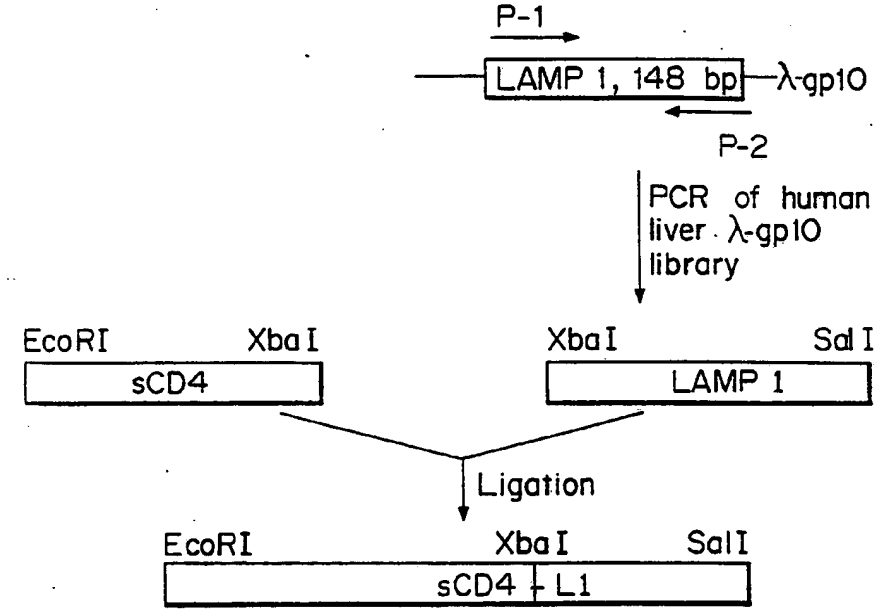
(72) Inventors: TANG, Jordan, J., N. ; 5600 N. Everest Avenue, Oklahoma City, OK 73111 (US). LIN, Xin-Li ; 824 N.W. 140th Street, Edmond, OK 73013 (US).

(74) Agents: PABST, Patrea, L. et al.; Kilpatrick & Cody, 1100 Peachtree Street, Suite 2800, Atlanta, GA 30309-4530 (US).

(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, SE).

Published
*With international search report.*
*Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.*

(54) Title: FUSION PROTEINS TARGETED TO LYSOSOMES, FOR THE TREATMENT OF AIDS




(57) Abstract

Fusion genes are constructed between a soluble CD4 (sCD4) gene, or portions thereof, and the genes of a lysosome targeting domain. Upon the biosynthesis of the fusion proteins in HIV infected cells, the sCD4 moiety binds newly synthesized gp160 and the lysosomal targeting moieties transport the entire complex to lysosomes. The genes diverts HIV coat glycoprotein to lysosomes for degradation, thus preventing the assembly of new virions and the propagation of HIV. The same process can be used for the treatment of other retroviruses.

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | |
|---|---|---|---|---|---|
| AT | Austria | FI | Finland | MN | Mongolia |
| AU | Australia | FR | France | MR | Mauritania |
| BB | Barbados | GA | Gabon | MW | Malawi |
| BE | Belgium | GB | United Kingdom | NL | Netherlands |
| BF | Burkina Faso | GN | Guinea | NO | Norway |
| BG | Bulgaria | GR | Greece | NZ | New Zealand |
| BJ | Benin | HU | Hungary | PL | Poland |
| BR | Brazil | IE | Ireland | PT | Portugal |
| CA | Canada | IT | Italy | RO | Romania |
| CF | Central African Republic | JP | Japan | RU | Russian Federation |
| CG | Congo | KP | Democratic People's Republic of Korea | SD | Sudan |
| CH | Switzerland | KR | Republic of Korea | SE | Sweden |
| CI | Côte d'Ivoire | LI | Liechtenstein | SK | Slovak Republic |
| CM | Cameroon | LK | Sri Lanka | SN | Senegal |
| CS | Czechoslovakia | LU | Luxembourg | SU | Soviet Union |
| CZ | Czech Republic | MC | Monaco | TD | Chad |
| DE | Germany | MG | Madagascar | TG | Togo |
| DK | Denmark | ML | Mali | UA | Ukraine |
| ES | Spain | | | US | United States of America |

# FUSION PROTEINS TARGETED TO LYSOSOMES, FOR THE TREATMENT OF AIDS

## Background of the Invention

In 1981, acquired immune deficiency syndrome (AIDS) was identified as a disease that severely compromises the human immune system, and that almost without exception leads to death. In 1983, the etiological cause of AIDS was determined to be the human immunodeficiency virus (HIV). In December, 1990, the World Health Organization estimated that between 8 and 10 million people worldwide were infected with HIV, and of that number, between 1,000,000 and 1,400,000 were in the U.S.

There are at least two types of HIV, Type I and Type II. Both preferentially infect T4 helper T lymphocytes and macrophages by interacting with the molecule CD4 on the surface of the target cell. All viruses infect cells by binding to the cell of an envelope protein. In the case of HIV, the envelope protein is gp120; the cell surface protein is an antigen called CD4. The viral membrane then apparently fuses with the cell membrane and the viral genes are injected into the cell, where they are replicated and new virions assembled using the host replicative processes. In some cases, the viral DNA most be integrated into the host genome, where it can remain latent for many years.

This replicative process has made it extremely difficult to treat, or as importantly, to cure HIV infection. Most attempts to vaccinate people against the disease have been unsuccessful; but at best would only limit infection. Most drugs have been targeted to replication of the viral nucleic acid.

In 1985, it was reported that the synthetic nucleoside 3'-azido-3'-deoxythymidine (AZT) inhibits the replication of human immunodeficiency virus type 1. Since then, a number of other synthetic nucleosides, including 2',3'-dideoxyinosine (DDI), 2',3'-dideoxycytidine (DDC), 3'-fluoro-3'-deoxythymidine (FLT), 2',3'-dideoxy,2',3'-didehydrothymidine (D4T), and 3'-azido-2',3'-dideoxyuridine (AZDU),

have been demonstrated to be effective against HIV, although none are able to cure the disease nor do more than prolong the life expectancy of the infected individuals.

The only drug commercially available for the treatment of AIDS, 3'-Azido-3'-deoxythymidine is a potent inhibitor of HIV reverse transcriptase. However, the benefits of AZT must be weighed against the severe adverse reactions of bone marrow suppression, nausea, myalgia, insomnia, severe headaches, anemia, peripheral neuropathy, and seizures. These adverse side effects often occur immediately after treatment begins, even though a minimum of six weeks of therapy is necessary to realize AZT's benefits. DDI, which has recently been approved by the FDA for the clinical testing for the treatment of AIDS, is also associated with a number of side effects, including sporadic pancreatis and peripheral neuropathy.

It is therefore apparent that there remains an important need to develop alternative therapies for treatment of HIV infections.

Gene therapy to achieve "intracellular immunization", as described by Baltimore, D. Nature 335: 395-396 (1988), against AIDS and other viral infections, especially of infections with retroviruses, offers definitive advantages because its successful application can potentially provide the patients with an intrinsic means to control the disease. To develop gene therapies for AIDS, it is equally important to generate new genes which can be used in AIDS therapy and to have the technology for gene transfer into the patients. Rapid advances in the technology for efficient gene transfer *in vitro* and *in vivo* have been made recently, for example, as reported by Friedmann, T. Science 244: 1275-1281 (1989). The developments of gene transfer vehicles, such as viral vectors, has led to clinical experimentation of gene therapy.

It is expected that the advances in gene therapy technology will continue. However, there are very few genes which have been

demonstrated to be effective against HIV and potentially useful for gene therapy against AIDS. HIV glycoprotein gp160 (precursor of gp120 and gp41) is synthesized on polysomes and transported through the secretory pathway [endoplasmic reticulum (ER), Golgi, and secretory vesicles] to the cell surface where the assembly of new virions takes place. The strong binding of viral gp120 to cell surface receptor CD4 is the primary route of HIV invasion of human cells (Klatzman, D., et al., Science 225: 59-63 (1984); Sattentau, Q.J., et al., Science 234: 1120-1123 (1986); Deen, K.C., et al., Nature 331: 82-84 (1988); Traunecher, A., et al., Nature 331: 84-86 (1988); Dalgleish, A.G., et al., Nature 312: 763-766 (1984); Maddon, P.J., et al., Cell 47: 333-348 (1986); Ho, D.D., et al., J. Clin. Invest. 77: 1712-1715 (1986); Gartner, S., et al., Science 233: 215-219 (1986)). The newly synthesized gp160 in HIV-infected T lymphocyte cells can bind the newly synthesized CD4 molecules in the ER (Hoxie, J.A., et al., Science 234: 1123-1127 (1986); and Kawamura, I., et al., J. Viol. 63: 3748-3754 (1989)).

Buonocore and Rose constructed a modified sCD4 with an addition to the C-terminus of a 6-residue sequence (Sequence ID No. 5), SEKDEL, which is the signal for ER retention (Munro, S., and Pelham, H.R.B. Cell 48: 899-907 (1987)). The modified CD4, sCD4-KDEL, stayed in the ER and prevented the newly synthesized gp160 from reaching the cell surface (Buonocore, L., and Rose, J.K. Nature 345: 625-628 (1990)). The ER residence of sCD4-KDEL has an clear limitation as a therapeutic agent, however. Proteins which reside in the ER, such as BiP, are transferred along the newly synthesized secretory proteins to the salvage compartment where the ER residence proteins are sorted and returned to ER (Pelham, H.R.B. Ann. Rev. Cell Biol. 5: 1-23 (1989)). For sCD4-KDEL to be effective as an anti-HIV agent, it must be synthesized continuously and at a level higher than that for gp160. This means that the continuing synthesis of sCD4-KDEL will

ultimately exceed the capacity of sorting mechanism in the salvage compartment. After that point, the newly synthesized sCD4-KDEL and gp160 will be lost from the ER, rendering sCD4-KDEL ineffective against HIV. It may also cause the loss of native ER residence proteins. Moreover, sCD4-KDEL was not shown to resist HIV infection or propagation.

It is therefore an object of the present invention to provide a gene therapy for use in treating or preventing AIDS and other retroviral diseases.

## Abstract of the Invention

This invention consists of the design and demonstration of fusion genes which can be used in the gene therapy for treating acquired immunodeficiency syndrome (AIDS) and other retroviruses. The principle of the therapeutic function is that upon transfer of these genes into human cells, the genes direct the synthesis of fusion proteins which interfere with the normal function of human immunodeficient virus (HIV), the causative agent of AIDS. The therapeutic genes are fusions between the genes encoding soluble CD4 (sCD4) (or other protein required for binding of virus to the target cell) and a lysosome targeting protein domain. Results have shown that when the fusion genes are expressed, the sCD4 moiety binds HIV glycoprotein gp160 in the endoplasmic reticulum while the lysosome targeting moiety transports the entire complex to the lysosomes for degradation. Thus the therapeutic genes prevent gp160 from reaching the cell surface, stopping the assembly of new virions and the propagation of HIV. The lysosome targeting domains successfully used in the fusion genes are human procathepsin D (PCaD) and parts of human lysosomal membrane proteins lamp-1, lamp-2, and acid phosphatase.

The experimental evidences that established the predicted function of the new genes include the following: (a) the transfection of gp160 gene into HeLa cells resulted in finding gp160 protein on the cell surface. The transfection of one of the therapeutic genes with gp160 genes into HeLa cells stopped gp160 to reach the cell surface; (b) in the presence of one of the therapeutic genes in HeLa cells, gp160 protein is degraded more rapidly than in the absence of the therapeutic gene; (c) the newly synthesized therapeutic proteins in HeLa cells are cleaved and digested in a manner characteristic of lysosomal activity; (d) transfecting one of the therapeutic genes into cultured T-lymphocyte cell line CEM inhibited the propagation of HIV.

## Brief Description of the Drawings

Figure 1 is a schematic presentation of the overall strategy of the synthesis and secretion of proteins in the endoplasmic reticulum (ER) and Golgi. In ER, gp160 of HIV is synthesized and glycosylated, and a fusion protein of soluble CD4-procathepsin D (sCD4-PCaD) is expressed from a cloned gene. The pro-cathepsin moiety of the sCD4-PCaD is N-glycosylated and mannose phosphorylated since it is a lysosomal enzyme. Then the gp160 binds to the sCD4 moiety of sCD4-PCaD. The complex is transported through the cis-Golgi to trans-Golgi. In trans-Golgi network, the mannose-6-phosphate (Man-$PO_4$) receptors bind mannose-6-phosphate of the PCaD moiety and target the whole complex, including gp160, to the lysosomes. In the lysosomes, procathepsin D activates to cathepsin D, and gp160 and sCD4 moieties are proteolytically degraded. This strategy is designed to prevent gp160 from entering the secretory pathway to reach the cell surface, thus preventing the assembly of HIV.

Figure 2 is a schematic of the construction of sCD4-PCaD. The PCR primers used are:

P-1, 5'-GAATTCAAGCCCAGAGCCCTGCC-3' (Sequence ID No. 6)

P-2, 5'-TCTAGAGGCCATTGGCTGCACCG-3' (Sequence ID No. 7)

P-3, 5'-TCTAGACTCGTCAGGATCCCGCTG-3' (Sequence ID No. 8)

P-4, 5'-GTCGACCTAGAGGCGGGCAGCC-3' (Sequence ID No. 9)

Figure 3 is a schematic of the construction of sCD4-HAP. The PCR primers used are:

P-1 = 5'-TCTAGACAGCTGGCAAGCGGTCCTG-3' (Sequence ID No. 10)

P-2 = 5'-GTCGACTCAGGCGTGGTCCTCCCC-3' (Sequence ID No. 11)

Figure 4 is a schematic of the construction of sCD4-L1. The PCR primers used are:

P-1 = 5'-TCTAGACTGCTGGACGAGAACAGCAC-3' (Sequence ID No. 12)

P-2 = 5'-GTCGACACCAGGCTAGATAGTCTGGTAG-3' (Sequence ID No. 13)

Figure 5 is a schematic of the construction of sCD4-L2. The PCR primers used are:

P-1 = 5'-TCTAGAAGTGCAGATGACGACAACTTC-3' (Sequence ID No. 14)

P-2 = 5'-GTCGACCTAAAATTGCTCATATCCAGCATG-3' (Sequence ID No. 15)

Figure 6 is a graph of reverse transcriptase (millions dpm/ml) versus time for blank vector pRc/RSV (1, -X-); no DNA (2, squares); pRc/RSV - sCD4-PCaD (3, inverted triangle); pRc/RSV - sCD4-HAP (4, -#-); pRc/RSV - sCD4-L2 (5, triangle); and pRc/RSV - sCD4-L2 (6, -diamond-).

Figures 7A and B are graphs of ß-hexosaminidase and density (g/ml) versus gradient fraction for Percoll gradients of lysosomes, showing the distribution of gp160 in fractions from Percoll density gradient centrifugation in the presence and absence of sCD4-PCaD gene expression. Figure 7A shows the β-hexosaminidase activity (solid line), density (broken line), and the autoradiography of the gel electrophoresis from cells transfected with gp160 alone. Figure 7B are the same data from cells cotransfected with gp160 and sCD4-PCaD genes.

## Detailed Description of the Invention

A method for treating retroviral infections, especially human immunodeficiency virus (HIV), wherein a fusion protein is created that binds to the viral envelope protein as it is formed and transports the envelope protein to lysosomes where it is degraded. The result is that the ability of the virus to replicate within the cell producing the fusion protein is limited. The fusion protein consists of two components: the protein which binds to the viral envelope protein and a protein (or domain of the protein) which targets the fusion protein to a lysosome. In the preferred embodiment for treating HIV, the first protein is soluble CD4, which binds HIV glycoprotein gp160, and the second protein is procathepsin D (PCaD), parts of human lysosomal membrane proteins lamp-1, lamp-2, or acid phosphatase. In the most preferred embodiment, the fusion protein is sCD4-L1 or sCD4-L2.

The teachings of the following references cited herein are specifically incorporated herein as exemplifying methods or reagents useful in constructing and using the fusion proteins for treatment of viral disorders.

The sCD4 fusion proteins which bind to gp160 and sort as a complex to lysosomes for degradation have the advantage over the ER retention shown by the modified sCD4 with an addition to the C-terminus

of a 6-residue sequence (Sequence ID No. 5), SEKDEL, which is the signal for ER retention, of Buonocore and Rose, because of the continuing removal of bound gp160 from the ER/Golgi system.

sCD4 in the fusion gene can be substituted by its parts, domain D1 or combined domains D1-D2. sCD4, the extracellular segment of CD4, consists of four tandem immunoglobulin-like domains. The N-terminal domain of sCD4, D1, by itself binds gp120 with high affinity, as reported by Arthos, J., et al. Cell 57, 469-481 (1989). Active recombinant D1 and D1-D2 domains have been obtained by (Arthos, et al.; Chao, B.H., et al., J. Biol. Chem. 264, 5812-5816 (1989); Traunecker, A., et al., Nature 331, 84-86 (1988); and Berger, E., et al. Proc. Natl. Acad. Sci. USA, 85: 2357-2361 (1988)), indicating that these domains are capable of independent folding.

Two types of lysosomal targeting components are particularly suited for the fusion of sCD4. First is a lysosomal proenzyme which contains a structural marker for lysosomal targeting. Procathepsin D (PCaD) was chosen as the lysosome targeting domain of the prototype therapeutic gene for the reason that much is known of its structure and function relationships (Tang, J. and Wong, R.N.S. J. Cell. Biochem. 33: 53-63 (1987); Takahashi, T., et al., J. Biol. Chem., 258: 2819-2830 (1983); Shewale, J.G., and Tang, J. Proc. Natl. Acad. Sci. USA, 80: 3703-3707 (1984); Yonezawa, S., et al., J. Biol. Chem. 263: 16504-16511 (1988); Faust, P.L., et al., Proc. Natl. Acad. Sci. USA, 82: 4910-4914 (1985)), its sorting mechanism via mannose-6-phosphate receptors in the trans-Golgi network (Kornfeld, S. and Mellman, I. Ann. Rev. Cell Biol. 5: 483-525 (1989)), and the spontaneous activation of its precursor, procathepsin D, in the lysosomes (Hasilick, A., et al., Eur. J. Biochem. 125: 37-321 (1982)).

The second type of lysosomal targeting domain for sCD4 fusion is taken from part of the lysosomal membrane proteins. Three human

lysosomal membrane proteins, lamp-1 (L1), lamp-2 (L2), and lysosomal acid phosphatase (HAP) (Fukuda, M., et al., J. Biol. Chem. 263: 18920-18928 (1988); Pohlmann, R. et al. EMBO J. 7: 2343-2350 (1988); and Waheed, A. et al. ibid. 7: 2351-2358 (1988)); each contains in its C-terminus a "lysosomal targeting signal" (LTS) region which consists of a short membrane-anchoring sequence and a short cytosolic domain (Kornfeld, S. and Mellman, I. Ann. Rev. Cell Biol. 5: 483-525 (1989); Peters, C., et al., EMBO J. 9: 3497-3506 (1990); William, M.A. and Fukuda, M. J. Cell Biol. 111: 955-966 (1990)). These three sCD4-LTS fusion genes, sCD4-HAP, sCD4-L1, and sCD4-L2, form the second group of therapeutic genes. The basic function of the sCD4-LTS therapeutic genes is the same as that in sCD4-PCaD, even though the targeting mechanisms are different between these two groups.

The PCaD moiety can be substituted for by other soluble lysosomal enzymes, such as other lysosomal soluble proteins containing mannose-6-phosphate markers. Some examples of human lysosomal enzymes with known cDNA structures and their cDNA sizes are α-N-Acetylgalactosaminidase, 1.3 kbp (Wang, A.M., et al., J. Biol. Chem. 265: 21859-21866 (1990)); Glycosylasparaginase, 1.1 kbp (Fisher, K.J., et al., FEBS Lett. 276: 440-444 (1990)); Glucocerebrosidase, 1.8 kbp (Tsuji, S., et al., J. Biol. Chem. 261: 50-53 (1986)); Procathepsin L, 1.1 kbp (Gal, S. and Gottesman, M.M. Biochem. J. 253: 303-306 (1988)); Procathepsin B, 1.1 kbp (Chan, S.J., et al., Proc. Natl. Acad. Sci. USA 83: 7721-7725 (1986)); and Procathepsin E, 1.2 kbp (Azuma, T., et al., J. Biol. Chem. 264: 16748-16753 (1986)). The lysosomal targeting domains (the transmembrane domains and cytosolic domains) of other lysosomal membrane proteins (human or other species origins) can also be substituted for human HAP, L1 and L2 in the therapeutic fusion genes.

It is believed that it will be possible to use more than one lysosome targeting domain in a therapeutic gene. Multiple targeting domains can increase the efficiency and capacity of the gp160 transporting to lysosomes. Examples of multidomain therapeutic genes are sCD4-PCaD-HAP, sCD4-PCaD-PCaD, and other combinations. The linker peptide between the sCD4 and the lysosome targeting domain can also be altered in amino acid sequence and in length to achieve different efficiencies and capacity of gp160 transporting to lysosomes.

The gene encoding the fusion protein is constructed using standard genetic engineering techniques, as described in detail below for treatment of HIV. The fusion gene is then introduced into cells *in vitro* using methods such as calcium phosphate coprecipitation, lipofection (liposomes), cell fusion, electroporation, or a vector such as a vaccinia virus. For example, CEM-SS cells were grown to exponential phase. Plasmids containing 10 $\mu$g each of the sCD4-LTD genes were separately transformed into 2 x $10^7$ cells by electroporation using the method of Aldovini, A. and M.B. Feinberg, Techniques in HIV Research eds. Aldovini, A. and B.D. Walker, pp. 147-176 (Stockton Press, NY 1990).

For *in vivo* applications, the preferred method is by transfection using a vector system such as the vaccinia virus. The vector is not limited however, for efficient expression, mammalian or viral promoters and other regulatory elements must be present. A recent account of the progress in gene therapy of humans is described in Friedmann, T., Science 244: 1275-1281 (1989). Some vector systems for introduction of therapeutic genes into AIDS patients have been described by Chimada, T., et al., J. Clin. Invest. 88:1043-1047.

The following examples demonstrate the effectiveness of the method for treating HIV infection. However, the same strategy could be used with other retroviral infections, such as hepatitis B and HTLV-1

caused leukemia. In these diseases, the cell surface protein receptors which bind virus glycoprotein can be used in fusion with lysosome targeting protein.

The present invention will be further understood by reference to the following non-limiting examples of the construction and cloning of therapeutic fusion genes, and their efficacy in degrading HIV envelope protein as it is formed.

**Example 1: Construction and Cloning of the Therapeutic Genes.**

Four sCD4-fusion genes, sCD4-PCaD, sCD4-HAP, sCD4-L1, and sCD4-12, were constructed as shown in the following sequences. Schematics of their constructions are shown in Figures 2-5, respectively. The gene encoding human procathepsin D cDNA was reported by Faust, et al., (1985). The gene encoding human lamp-1 and lamp-2 was reported by Fukuda, et al., (198). The gene encoding human HAP was reported by Pohlmann, et al., (1988).

The nucleotide sequence encoding sCD4-PCaD is shown below in Sequence ID No. 1. Underlined letters are engineered restriction sites.

GAATTCAAGC CCAGAGCCCT GCCATTTCTG TGGGCTCAGG TCCCTACTGC TCAGCCCCTT
CCTCCCTCGG CAAGGCCACA ATGAACCGGG GAGTCCCTTT TAGGCACTTG CTTCTGGTGC
TGCAACTGGC GCTCCTCCCA GCAGCCACTC AGGGAAAGAA AGTGGTGCTG GGCAAAAAAG
GGGATACAGT GGAACTGACC TGTACAGCTT CCCAGAAGAA GAGCATACAA TTCCACTGGA
AAAACTCCAA CCAGATAAAG ATTCTGGGAA ATCAGGGCTC CTTCTTAACT AAAGGTCCAT
CCAAGCTGAA TGATCGCGCT GACTCAAGAA GAAGCCTTTG GGACCAAGGA AACTTCCCCC
TGATCATCAA GAATCTTAAG ATAGAAGACT CAGATACTTA CATCTGTGAA GTGGAGGACC
AGAAGGAGGA GGTGCAATTG CTAGTGTTCG GATTGACTGC CAACTCTGAC ACCCACCTGC
TTCAGGGGCA GAGCCTGACC CTGACCTTGG AGAGCCCCCC TGGTAGTAGC CCCTCAGTGC
AATGTAGGAG TCCAAGGGGT AAAAACATAC AGGGGGGGAA GACCCTCTCC GTGTCTCAGC
TGGAGCTCCA GGATAGTGGC ACCTGGACAT GCACTGTCTT GCAGAACCAG AAGAAGGTGG
AGTTCAAAAT AGACATCGTG GTGCTAGCTT TCCAGAAGGC CTCCAGCATA GTCTATAAGA
AAGAGGGGGA ACAGGTGGAG TTCTCCTTCC CACTCGCCTT TACAGTTGAA AAGCTGACGG
GCAGTGGCGA GCTGTGGTGG CAGGCGGAGA GGGCTTCCTC CTCCAAGTCT TGGATCACCT
TTGACCTGAA GAACAAGGAA GTGTCTGTAA AACGGGTTAC CCAGGACCCT AAGCTCCAGA

TGGGCAAGAA GCTCCCGCTC CACCTCACCC TGCCCCAGGC CTTGCCTCAG TATGCTGGCT
CTGGAAACCT CACCCTGGCC CTTGAAGCGA AAACAGGAAA GTTGCATCAG GAAGTGAACC
TGGTGGTGAT GAGAGCCACT CAGCTCCAGA AAAATTTGAC CTGTGAGGTG TGGGGACCCA
CCTCCCCTAA GCTGATGCTG AGCTTGAAAC TGGAGAACAA GGAGGCAAAG GTCTCGAAGC
GGGAGAAGGC GGTGTGGGTG CTGAACCCTG AGGCGGGGAT GTGGCAGTGT CTGCTGAGTG
ACTCGGGACA GGTCCTGCTG GAATCCAACA TCAAGGTTCT GCCCACATGG TCCACCCCGG
TGCAGCCAAT GGCCTCTAGA CTCGTCAGGA TCCCGCTGCA CAAGTTCACG TCCATCCGCC
GGACCATGTC GGAGGTTGGG GGCTCTGTGG AGGACCTGAT TGCCAAAGGC CCCGTCTCAA
AGTACTCCCA GGCGGTGCCA GCCGTGACCG AGGGGCCCAT TCCCGAGGTG CTCAAGAACT
ACATGGACGC CCAGTACTAC GGGGAGATTG GCATCGGGAC GCCCCCCCAG TGCTTCACAG
TCGTCTTCGA CACGGGCTCC TCCAACCTGT GGGTCCCCTC CATCCACTGC AAACTGCTGG
ACATCGCTTG CTGGATCCAC CACAAGTACA ACAGCGACAA GTCCAGCACC TACGTGAAGA
ATGGTACCTC GTTTGACATC CACTATGGCT CGGGCAGCCT CTCCGGGTAC CTGAGCCAGG
ACACTGTGTC GGTGCCCTGC CAGTCAGCGT CGTCAGCCTC TGCCCTGGGC GGTGTCAAAG
TGGAGAGGCA GGTCTTTGGG GAGGCCACCA AGCAGCCAGG CATCACCTTC ATCGCAGCCA
AGTTCGATGG CATCCTGGGC ATGGCCTACC CCCGCATCTC CGTCAACAAC GTGCTGCCCG
TCTTCGACAA CCTGATGCAG CAGAAGCTGG TGGACCAGAA CATCTTCTCC TTCTACCTGA
GCAGGGACCC AGATGCGCAG CCTGGGGGTG AGCTGATGCT GGGTGGCACA GACTCCAAGT
ATTACAAGGG TTCTCTGTCC TACCTGAATG TCACCCGCAA GGCCTACTGG CAGGTCCACC
TGGACACAGG CACTTCCCTC ATGGTGGGCC CGGTGGATGA GGTGCGCGAG CTGCAGAAGG
CCATCGGGGC CGTGCCGCTG ATTCAGGGCG AGTACATGAT CCCCTGTGAG AAGGTGTCCA
CCCTGCCCGC GATCACACTG AAGCTGGGAG GCAAAGGCTA CAAGCTGTCC CCAGAGGACT
ACACGCTCAA GGTGTCGCAG GCCGGGAAGA CCCTCTGCCT GAGCGGCTTC ATGGGCATGG
ACATCCCGCC ACCCAGCGGG CCACTCTGGA TCCTGGGCGA CGTCTTCATC GGCCGCTACT
ACACTGTGTT TGACCGTGAC AACAACAGGG TGGGCTTCGC CGAGGCTGCC CGCCTCTAGC
AGCTG

The nucleotide sequence encoding sCD4-HAP is shown below in Sequence ID No. 2. Underlined letters are engineered restriction sites.

GAATTCAAGC CCAGAGCCCT GCCATTTCTG TGGGCTCAGG TCCCTACTGC TCAGCCCCTT
CCTCCCTCGG CAAGGCCACA ATGAACCGGG GAGTCCCTTT TAGGCACTTG CTTCTGGTGC
TGCAACTGGC GCTCCTCCCA GCAGCCACTC AGGGAAAGAA AGTGGTGCTG GGCAAAAAAG
GGGATACAGT GGAACTGACC TGTACAGCTT CCCAGAAGAA GAGCATACAA TTCCACTGGA
AAAACTCCAA CCAGATAAAG ATTCTGGGAA ATCAGGGCTC CTTCTTAACT AAAGGTCCAT
CCAAGCTGAA TGATCGCGCT GACTCAAGAA GAAGCCTTTG GGACCAAGGA AACTTCCCCC
TGATCATCAA GAATCTTAAG ATAGAAGACT CAGATACTTA CATCTGTGAA GTGGAGGACC

AGAAGGAGGA GGTGCAATTG CTAGTGTTCG GATTGACTGC CAACTCTGAC ACCCACCTGC
TTCAGGGGCA GAGCCTGACC CTGACCTTGG AGAGCCCCCC TGGTAGTAGC CCCTCAGTGC
AATGTAGGAG TCCAAGGGGT AAAAACATAC AGGGGGGGAA GACCCTCTCC GTGTCTCAGC
TGGAGCTCCA GGATAGTGGC ACCTGGACAT GCACTGTCTT GCAGAACCAG AAGAAGGTGG
AGTTCAAAAT AGACATCGTG GTGCTAGCTT TCCAGAAGGC CTCCAGCATA GTCTATAAGA
AAGAGGGGGA ACAGGTGGAG TTCTCCTTCC CACTCGCCTT TACAGTTGAA AAGCTGACGG
GCAGTGGCGA GCTGTGGTGG CAGGCGGAGA GGGCTTCCTC CTCCAAGTCT TGGATCACCT
TTGACCTGAA GAACAAGGAA GTGTCTGTAA AACGGGTTAC CCAGGACCCT AAGCTCCAGA
TGGGCAAGAA GCTCCCGCTC CACCTCACCC TGCCCCAGGC CTTGCCTCAG TATGCTGGCT
CTGGAAACCT CACCCTGGCC CTTGAAGCGA AAACAGGAAA GTTGCATCAG GAAGTGAACC
TGGTGGTGAT GAGAGCCACT CAGCTCCAGA AAAATTTGAC CTGTGAGGTG TGGGGACCCA
CCTCCCCTAA GCTGATGCTG AGCTTGAAAC TGGAGAACAA GGAGGCAAAG GTCTCGAAGC
GGGAGAAGGC GGTGTGGGTG CTGAACCCTG AGGCGGGGAT GTGGCAGTGT CTGCTGAGTG
ACTCGGGACA GGTCCTGCTG GAATCCAACA TCAAGGTTCT GCCCACATGG TCCACCCCGG
TGCAGCCAAT GGCCTCTAGA CAGCTGGCAA GCGGTCCTGC AGACACAGAG GTGATTGTGG
CCTTGGCTGT ATGTGGCTCC ATCCTCTTCC TCCTCATAGT GCTGCTCCTC ACCGTCCTCT
TCCGGATGCA GGCCCAGCCT CCTGGCTACC GCCACGTCGC AGATGGGGAG GACCACGCCT
GAGTCGAC

The nucleotide sequence encoding sCD4-L1 is shown below in Sequence ID No. 3. Underlined letters are engineered restriction sites.

GAATTCAAGC CCAGAGCCCT GCCATTTCTG TGGGCTCAGG TCCCTACTGC TCAGCCCCTT
CCTCCCTCGG CAAGGCCACA ATGAACCGGG GAGTCCCTTT TAGGCACTTG CTTCTGGTGC
TGCAACTGGC GCTCCTCCCA GCAGCCACTC AGGGAAAGAA AGTGGTGCTG GGCAAAAAAG
GGGATACAGT GGAACTGACC TGTACAGCTT CCCAGAAGAA GAGCATACAA TTCCACTGGA
AAAACTCCAA CCAGATAAAG ATTCTGGGAA ATCAGGGCTC CTTCTTAACT AAAGGTCCAT
CCAAGCTGAA TGATCGCGCT GACTCAAGAA GAAGCCTTTG GGACCAAGGA AACTTCCCCC
TGATCATCAA GAATCTTAAG ATAGAAGACT CAGATACTTA CATCTGTGAA GTGGAGGACC
AGAAGGAGGA GGTGCAATTG CTAGTGTTCG GATTGACTGC CAACTCTGAC ACCCACCTGC
TTCAGGGGCA GAGCCTGACC CTGACCTTGG AGAGCCCCCC TGGTAGTAGC CCCTCAGTGC
AATGTAGGAG TCCAAGGGGT AAAAACATAC AGGGGGGGAA GACCCTCTCC GTGTCTCAGC
TGGAGCTCCA GGATAGTGGC ACCTGGACAT GCACTGTCTT GCAGAACCAG AAGAAGGTGG
AGTTCAAAAT AGACATCGTG GTGCTAGCTT TCCAGAAGGC CTCCAGCATA GTCTATAAGA
AAGAGGGGGA ACAGGTGGAG TTCTCCTTCC CACTCGCCTT TACAGTTGAA AAGCTGACGG
GCAGTGGCGA GCTGTGGTGG CAGGCGGAGA GGGCTTCCTC CTCCAAGTCT TGGATCACCT
TTGACCTGAA GAACAAGGAA GTGTCTGTAA AACGGGTTAC CCAGGACCCT AAGCTCCAGA

TGGGCAAGAA GCTCCCGCTC CACCTCACCC TGCCCCAGGC CTTGCCTCAG TATGCTGGCT
CTGGAAACCT CACCCTGGCC CTTGAAGCGA AAACAGGAAA GTTGCATCAG GAAGTGAACC
TGGTGGTGAT GAGAGCCACT CAGCTCCAGA AAAATTTGAC CTGTGAGGTG TGGGGACCCA
CCTCCCCTAA GCTGATGCTG AGCTTGAAAC TGGAGAACAA GGAGGCAAAG GTCTCGAAGC
GGGAGAAGGC GGTGTGGGTG CTGAACCCTG AGGCGGGGAT GTGGCAGTGT CTGCTGAGTG
ACTCGGGACA GGTCCTGCTG GAATCCAACA TCAAGGTTCT GCCCACATGG TCCACCCCGG
TGCAGCCAAT GGCCTCTAGA CTGCTGGACG AGAACAGCAC GCTGATCCCC ATCGCTGTGG
GTGGTGCCCT GGCGGGGCTG GTCCTCATCG TCCTCATCGC CTACCTCGTC GGCAGGAAGA
GGAGTCACGC AGGCTACCAG ACTATCTAGC CTGGTGTCGA C

The nucleotide sequence encoding sCD4-L2 is shown below in Sequence ID No. 4. Underlined letters are engineered restriction sites.

GAATTCAAGC CCAGAGCCCT GCCATTTCTG TGGGCTCAGG TCCCTACTGC TCAGCCCCTT
CCTCCCTCGG CAAGGCCACA ATGAACCGGG GAGTCCCTTT TAGGCACTTG CTTCTGGTGC
TGCAACTGGC GCTCCTCCCA GCAGCCACTC AGGGAAAGAA AGTGGTGCTG GGCAAAAAAG
GGGATACAGT GGAACTGACC TGTACAGCTT CCCAGAAGAA GAGCATACAA TTCCACTGGA
AAAACTCCAA CCAGATAAAG ATTCTGGGAA ATCAGGGCTC CTTCTTAACT AAAGGTCCAT
CCAAGCTGAA TGATCGCGCT GACTCAAGAA GAAGCCTTTG GGACCAAGGA AACTTCCCCC
TGATCATCAA GAATCTTAAG ATAGAAGACT CAGATACTTA CATCTGTGAA GTGGAGGACC
AGAAGGAGGA GGTGCAATTG CTAGTGTTCG GATTGACTGC CAACTCTGAC ACCCACCTGC
TTCAGGGGCA GAGCCTGACC CTGACCTTGG AGAGCCCCCC TGGTAGTAGC CCCTCAGTGC
AATGTAGGAG TCCAAGGGGT AAAAACATAC AGGGGGGGAA GACCCTCTCC GTGTCTCAGC
TGGAGCTCCA GGATAGTGGC ACCTGGACAT GCACTGTCTT GCAGAACCAG AAGAAGGTGG
AGTTCAAAAT AGACATCGTG GTGCTAGCTT TCCAGAAGGC CTCCAGCATA GTCTATAAGA
AAGAGGGGGA ACAGGTGGAG TTCTCCTTCC CACTCGCCTT TACAGTTGAA AAGCTGACGG
GCAGTGGCGA GCTGTGGTGG CAGGCGGAGA GGGCTTCCTC CTCCAAGTCT TGGATCACCT
TTGACCTGAA GAACAAGGAA GTGTCTGTAA AACGGGTTAC CCAGGACCCT AAGCTCCAGA
TGGGCAAGAA GCTCCCGCTC CACCTCACCC TGCCCCAGGC CTTGCCTCAG TATGCTGGCT
CTGGAAACCT CACCCTGGCC CTTGAAGCGA AAACAGGAAA GTTGCATCAG GAAGTGAACC
TGGTGGTGAT GAGAGCCACT CAGCTCCAGA AAAATTTGAC CTGTGAGGTG TGGGGACCCA
CCTCCCCTAA GCTGATGCTG AGCTTGAAAC TGGAGAACAA GGAGGCAAAG GTCTCGAAGC
GGGAGAAGGC GGTGTGGGTG CTGAACCCTG AGGCGGGGAT GTGGCAGTGT CTGCTGAGTG
ACTCGGGACA GGTCCTGCTG GAATCCAACA TCAAGGTTCT GCCCACATGG TCCACCCCGG
TGCAGCCAAT GGCCTCTAGA AGTGCAGATG ACGACAACTT CCTTGTGCCC ATAGCGGTGG

GAGCTGCCTT GGCAGGAGTA CTTATTCTAG TGTTGCTGGC TTATTTTATT GGTCTCAAGC
ACCATCATGC TGGATATGAG CAATTTTAGG TCGAC

**Example 2: Demonstration of efficacy of the Therapeutic Genes:**

**I. Hela-CD4+ Cells Expressing Gp160 Formed Syncytia which was Reversed by Coexpressing of Therapeutic Genes.**

HeLa-CD4+ cell are HeLa cells which express CD4 on the cell surface. When the gp160 gene is transfected into HeLa-CD4+ cells, the cells have both CD4 and gp160 on the cell surface. Because CD4 binds gp160 tightly, the cells aggregate and fuse together to form giant cell masses called synthetium. When gp160 and a therapeutic gene are both transfected into the cells, even though the gp160 is synthesized, it does not appear on the cell surface because the protein made from the therapeutic gene transports the gp160 to the lysosomes instead of the cell surface, so no syncytium is formed.

Syncytium formation was monitored for HeLa-CD4+ cells (clone HT4-6C) transfected with gp160 or cotransfected with gp160 and other genes, using a vaccinia virus expression system. The following results were obtained.

Untransfected cells ..... No syncytium

Cells transfected with procathepsin D (control) ..... No syncytium

Cells transfected with gp160 gene ..... Syncytia

Cells cotransfected with gp160 and sCD4-PCaD gene in the wrong cloning direction ..... Syncytia

Cells cotransfected with gp160 and blank pET-3a vector .....Syncytia

Cells cotransfected with gp160 and therapeutic gene sCD4-PCaD (right direction) .....No syncytium

Cells cotransfected with gp160 and therapeutic gene sCD4-HAP ..... No syncytium

Cells cotransfected with gp160 and sCD4 ..... Syncytia

Cells cotransfected with gp160 and sCD4-L1 ..... No syncytium

Cells cotransfected with gp160 and sCD4-L2 ..... No syncytium

These results were completely reproducible in three experiments using HeLa-CD4+ clone 6C and in one experiment with clone 1022. These results indicate that neither sCD4 nor the control vectors could reverse the syncytium formation. Only the therapeutic genes sCD4-PCaD and sCD4-HAP can reverse the syncytium formation by gp160. The expression of the lysosome targeting domain with the sCD4 is necessary to prevent the syncytium.

**II. Gp160 and the Therapeutic Gene Products (sCD4-PCaD and sCD4-HAP proteins) and gp160 are Degraded in the Lysosomes.**

The following reagents used in these studies were obtained through the AIDS Research and Reference Program, AIDS Program, NIAID, NIH: plasmid pIIIenv3-1 (reagent #289, from Dr. Joseph Sodroski), plasmid pt4B (#157, from Dr. Richard Axel), antiserum to HIV-1 gp120 (#288, from Dr. Michael Phelan), antiserum to CD4 (#314, from Dr. Michael Phelan), vector vTF7-3 (#356, from Drs. Tom Fuerst and Bernard Moss), HeLa cells (#153, from Dr. Richard Axel), HeLa-CD4+ (clone 6C) (#459, from Dr. Bruce Chesebro), HeLa-CD4+ (clone 1022) (#1109, from Dr. Bruce Chesebro), and CEM-SS cells (#776, from Dr. Peter L. Nara). Rabbit anticathepsin D antiserum was produced as described by Huang, J.S. et al., J. Biol. Chem. 254, 11405-11417 (1979).

HeLa cells were transfected with gp160 gene or cotransfected with gp160 and one of the therapeutic genes (sCD4-PCaD or sCD4-HAP), using a vaccinia expression system for transfection and expression. The proteins were metabolically labelled with radioactive methionine (pulse). At different times (chase), the cells were separated from the culture medium and homogenized. Immunoprecipitations were performed on cell homogenates and media with antibodies against gp160, CD4, and PCaD, respectively. The precipitated proteins were separated on SDS-gel electrophoresis and their patterns visualized by exposure to photographic films.

The pattern of the specific proteins indicate the fates of various gp160 and the therapeutic proteins. The results from the pulse-chase experiments indicate that gp160 is not degraded when the gene for gp160 is transfected alone. When gp160 and the therapeutic genes are cotransfected, both proteins are rapidly degraded. The results are summarized below.

The pattern of protein bands immunoprecipitated with gp160 antibody after a 4 h chase when gp160 was transfected alone, showed that both gp160 and gp120 bands were associated with the cells; but only gp120 was in the medium due to shedding from cell surface. Cotransfection of gp160 and sCD4 did not significantly change from gp160 alone. When gp160 was cotransfected with either sCD4-PCaD gene (sCD4-P) or sCD4-HAP gene (sCD4-H), no gp120 was observed either in the cells or in the media, indicating that gp160 did not enter into the pathway to the cell surface which also processes gp160 to gp120 and gp41. A band was observed at near 96 kD of the cells with sCD4-PCaD cotransfection. This band represents metabolically radiolabeled sCD4-PCaD which binds strongly to gp160 and coprecipitates by antibody against gp160. Similarly, a 50-kD band from the cells cotransfected with sCD4-HAP gene represents the fusion protein sCD4-

HAP. These results have been completely reproducible in several experiments.

The results demonstrate that sCD4-fusion genes cause gp160 to go to a different transporting pathway. Strong bindings between sCD4-fusion proteins and gp160 takes place in the cells.

The patterns from immunoprecipitation of cell homogenates and media with antibodies against gp160 obtained by SDS-gel electrophoresis also demonstrate that gp160 is degraded more rapidly in the presence of sCD4-PCaD gene. After a 4 h chase, the combined amount of gp160 and gp120 in the cells transfected with only gp160 gene already far exceeds the gp160 band from the cells cotransfected with gp160 and sCD4-PCaD genes.

After an 18 h chase, the amount of combined gp160 and gp120 (gp160 transfection) did not appreciably change from the 4 h result. However, the gp160 band is largely diminished in the cells cotransfected with the sCD4-PCaD fusion gene. (The intensity of this gp160 band is only about 1/8 of that of the combined gp160 + gp120.) These results indicate gp160 in the cells synthesizing sCD4-PCaD is rapidly degraded as a result of the expression of the sCD4-PCaD gene in the cells. This conclusion is consistent with the explanation that gp160 is transported by the fusion protein to lysosomes and degraded.

Cathepsin D antibody immunoprecipitates show similar results by SDS gel electrophoresis. At the end of a pulse or after a 2 h chase, a major band near 97 kD is clearly the fusion protein sCD4-PCaD. The molecular size of this band agrees with the fusion between sCD4 (apparent size: 46 kD) and procathepsin D (apparent size: 50 kD). With a 4 h chase, cathepsin D appeared mainly as a 35 kD band. This is close to the size of the human cathepsin D heavy chain, usually in the range of 30 to 35 kD. (The 0.9 kD light chain is usually too faint to be seen.) It is known that procathepsin is rapidly activated to cathepsin D

and slowly processed to a 2-chain enzyme in the lysosomes, as reported by Hasilik, A. and Neufeld, E.F. J. Biol. Chem. 255: 4937-4945 (1980) and Erickson, A.H., et al., J. Biol. Chem. 256: 11224-11231 (1981). Thus, the time for the fusion protein to reach lysosomes is between 2 to 4 h. In summary, the sCD4-PCaD fusion protein is targeted to the lysosomes, activated, and correctly processed.

The electrophoretic patterns of CD4 antibody immunoprecipitates at different chase times also show similar results. At the end of pulse labelling of 20 min (0 chase time), gp160 transfection alone provides only faint background bands, as no CD4 synthesis is expected in HeLa cells. Cotransfection with sCD4-PCaD produces a major band near 97 kD. This is the same position for the synthetic band detected with cathepsin D antibody. Even as early as at the end of the 20 min pulse period, some degradation bands of sCD4 could be seen, with the 46 kD band appearing at the same position as the authentic sCD4; so this band must have come from the activation of the procathepsin D moiety of the fusion protein in the lysosomes which resulted in the separation of sCD4 from PCaD. The major band for the cotransfection of sCD4-HAP of 0 time chase is about 50 kD. Since the lysosome targeting domain from HAP is only 45 amino acid residues, this is the expected size for the fusion protein sCD4-HAP.

After a 2 h chase, much of the labelled sCD4- domain in sCD4-PCaD had been degraded to 3 bands in the size range of 25 to 40 kD. The sCD4-HAP band had become a doublet due to the degradation of the targeting domain from the C-terminal of sCD4, which is known to occur for all the lysosomal membrane proteins. After 4 h and 16 h chase periods, diminishing amounts of material recognizable by CD4 antibody were present due to further lysosomal proteolysis.

In conclusion, some sCD4-PCaD fusion protein reaches lysosomes as early as 20 min. At the end of 2 h, the degradation of sCD4-fusion

proteins is extensive. At 16 h, the degradation is nearly complete. The degradation kinetics of sCD4- moiety is similar to that of gp160 which was nearly completed after 18 h.

**III. T-lymphocyte CEM cells transfected with therapeutic genes can resist HIV propagation.**

CEM(CD4+) cells are T-lymphoma cells which have CD4 on their cell surface. These cells are susceptible for HIV infection. The cells are transfected with therapeutic genes which transiently expressing therapeutic proteins. The transfected cells, control cells transfected with blank vectors, and untransfected cells are challenged with HIV. HIV propagation was determined by the analysis of reverse transcriptase activity.

As shown in Figure 6, reverse transcriptase (RT) activities at different time in control cells (electroporation without DNA), in cells separately transfected with blank vector, and vectors with one of four therapeutic genes (sCD4-PCaD, sCD4-HAP, sCD4-L1 and sCD4-L2). Transfections were done 40 h before the addition of HIV (day 0). Up to day 3, no difference could be seen in four cell groups. However, at day 7, the control cells and the cells transfected with blank vector both showed a dramatic increase in RT activity. The agreements among these two controls are reasonably good. In the cells transfected with sCD4-fusion genes, however, the increases are only about 10-15% of the controls.

The transient expression of the fusion genes usually peaks around two to three days and lasted for several days. However, the differences in RT activity became apparent at day 7. This means that from day 0 to day 3 the sCD4-fusion genes are depriving the HIV virions of gp160 (gp120); thus, the deficiency of gp160 on the virions was manifested as inhibition of propagation at day 7. These data support the view that the fusion genes work as intended in CEM cells challenged with HIV.

The results lead to the following conclusions:

Therapeutic genes can reverse syncytium formations. Thus, the therapeutic genes prevent gp160 from reaching cell surface. When the sCD4-fusion genes are expressed, gp160 and sCD4 are degraded rapidly. All evidence shows the main site of degradation is the lysosomes. When sCD4-PCaD is expressed, procathepsin D moiety is activated and correctly processed, which is exclusively a lysosomal phenomenon. Thus, sCD4-PCaD protein enters lysosomes. In summary,, it can be concluded that the therapeutic genes transport gp160 to lysosomes. Therapeutic genes sCD4-PCaD, sCD4-HAP, sCD4-L1, and sCD4-L2 inhibit HIV propagation in T-lymphocyte cell line CEM cells.

**Example 3: Expression of sCD4-PCaD gene causes gp160 to be present in the lysosomes.**

To further demonstrate sCD4-fusion proteins are degraded in the lysosomes and that gp160 is transported to the lysosomes by each of the 4 sCD4 fusion proteins, lysosomes were fractionated in Percoll density gradient centrifugation from the $^{35}$S-methionine labelled cells which had been either transfected with gp160 or cotransfected with gp160 and sCD4-PCaD genes. HeLa cells were transfected with gp160 gene or cotransfected with gp160 + sCD4-PCaD genes, labelled with $^{35}$S-methionine, and chased for 4 h as described in the pulse-chase experiments. Cells were scraped from plates, homogenized, and centrifuged to obtain the postnuclear supernate as described by Gieselmann, J. Cell Biol. 97: 1-5 (1983). This supernate (1 ml) was applied as a top layer of a 15% Percoll (Pharmacia) solution of 0.25 M sucrose in a Beckman polycarbonate centrifuge bottle (No. 355603) and centrifuged in a Beckman 80Ti rotor at 33,000 x g for 30 min. Fractions of 0.75 ml each, which were collected starting from the bottom, were analyzed for $\beta$-hexosaminidase activity (Geiger and Arnon, Methods in Enzymology 50, 547-555 (1978) and were subjected to

immunoprecipitation using anti-gp160 antiserum. The electrophoresis of precipitate and autoradiography are the same as described above. Figure 7A shows the β-hexosaminidase activity (solid line), density (broken line), and the autoradiography of the gel electrophoresis from cells transfected with gp160 alone. Figure 7B are the same data from cells cotransfected with gp160 and sCD4-PCaD genes.

The fractions were analyzed for lysosomes by β-hexosaminidase activity and also immunoprecipitated with anti-gp160 antiserum, electrophoresed, and visualized by autoradiography. Figures 7A and B shows that the activity of hexoaminidase accumulates in fractions 1-3, which represent dense lysosomes, and is more pronounced in fractions 10-13, which represent the light lysosomes and endosomes. The total β-hexosaminidase activates are similar in two transfections (Figures 7A and B), meaning that the number of lysosomes are about the same in them. However, gp160 bands are present only in the cells cotransfected with sCD4-PCaD gene, not in the cells transfected with gp160 alone. These observations support the view that the expression of sCD4-PCaD gene diverts gp160 to lysosomes.

**Example 4: Demonstration that the fusion gene can encode portions of the binding protein and be efficacious.**

D1 and D1-D2 were tested as alternatives of the sCD4 domain in the sCD4-fusion genes. The results from syncytium formation and pulse-chase experiments indicated that sCD4 in the fusion gene can be substituted by its parts, domain D1 or combined domains D1-D2.

The ability of D1-HAP and D1-D2-HAP to reverse the syncytium formation caused by the transfection of gp160 into HeLa-CD4+ cells. The experimental conditions are the same as already described for the other syncytium experiments. It was observed that both fusion genes, when cotransfected with gp160 gene, prevented the syncytium formation.

Pulse-chase studies using the fusion genes were also conducted in HeLa cells. Antiserum against gp160 was used to immunoprecipitate. It was observed that D1-HAP and D1-D2-HAP coprecipitated with gp160.

These observations suggest that D1 and D1-D2 fusion genes work as effectively as the sCD4 fusion genes.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Tang, J. N.

(ii) TITLE OF INVENTION: Fusion Protein Genes for Treatment of Aids

(iii) NUMBER OF SEQUENCES: 15

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Kilpatrick & Cody
(B) STREET: 1100 Peachtree Street
(C) CITY: Atlanta
(D) STATE: Georgia
(E) COUNTRY: U.S.
(F) ZIP: 30309-4530

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: US
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Pabst, Patrea L.
(B) REGISTRATION NUMBER: 31,284
(C) REFERENCE/DOCKET NUMBER: OMRF129

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 404-815-6508
(B) TELEFAX: 404-815-6555

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2465 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapien
(F) TISSUE TYPE: Epithelial

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..6
(D) OTHER INFORMATION: /note= "Restriction site"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 2460..2465
(D) OTHER INFORMATION: /note= "Restriction site"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GAATTCAAGC CCAGAGCCCT GCCATTTCTG TGGGCTCAGG TCCCTACTGC TCAGCCCCTT 60

CCTCCCTCGG CAAGGCCACA ATGAACCGGG GAGTCCCTTT TAGGCACTTG CTTCTGGTGC 120

TGCAACTGGC GCTCCTCCCA GCAGCCACTC AGGGAAAGAA AGTGGTGCTG GGCAAAAAAG 180

GGGATACAGT GGAACTGACC TGTACAGCTT CCCAGAAGAA GAGCATACAA TTCCACTGGA 240

AAAACTCCAA CCAGATAAAG ATTCTGGGAA ATCAGGGCTC CTTCTTAACT AAAGGTCCAT 300

CCAAGCTGAA TGATCGCGCT GACTCAAGAA GAAGCCTTTG GGACCAAGGA AACTTCCCCC 360

TGATCATCAA GAATCTTAAG ATAGAAGACT CAGATACTTA CATCTGTGAA GTGGAGGACC 420

AGAAGGAGGA GGTGCAATTG CTAGTGTTCG GATTGACTGC CAACTCTGAC ACCCACCTGC 480

TTCAGGGGCA GAGCCTGACC CTGACCTTGG AGAGCCCCCC TGGTAGTAGC CCCTCAGTGC 540

AATGTAGGAG TCCAAGGGGT AAAAACATAC AGGGGGGGAA GACCCTCTCC GTGTCTCAGC 600

TGGAGCTCCA GGATAGTGGC ACCTGGACAT GCACTGTCTT GCAGAACCAG AAGAAGGTGG 660

AGTTCAAAAT AGACATCGTG GTGCTAGCTT TCCAGAAGGC CTCCAGCATA GTCTATAAGA 720

AAGAGGGGGA ACAGGTGGAG TTCTCCTTCC CACTCGCCTT TACAGTTGAA AAGCTGACGG 780

GCAGTGGCGA GCTGTGGTGG CAGGCGGAGA GGGCTTCCTC CTCCAAGTCT TGGATCACCT 840

TTGACCTGAA GAACAAGGAA GTGTCTGTAA AACGGGTTAC CCAGGACCCT AAGCTCCAGA 900

TGGGCAAGAA GCTCCCGCTC CACCTCACCC TGCCCCAGGC CTTGCCTCAG TATGCTGGCT 960

CTGGAAACCT CACCCTGGCC CTTGAAGCGA AAACAGGAAA GTTGCATCAG GAAGTGAACC 1020

TGGTGGTGAT GAGAGCCACT CAGCTCCAGA AAAATTTGAC CTGTGAGGTG TGGGGACCCA 1080

CCTCCCCTAA GCTGATGCTG AGCTTGAAAC TGGAGAACAA GGAGGCAAAG GTCTCGAAGC 1140

GGGAGAAGGC GGTGTGGGTG CTGAACCCTG AGGCGGGGAT GTGGCAGTGT CTGCTGAGTG 1200

ACTCGGGACA GGTCCTGCTG GAATCCAACA TCAAGGTTCT GCCCACATGG TCCACCCCGG 1260

TGCAGCCAAT GGCCTCTAGA CTCGTCAGGA TCCCGCTGCA CAAGTTCACG TCCATCCGCC 1320

GGACCATGTC GGAGGTTGGG GGCTCTGTGG AGGACCTGAT TGCCAAAGGC CCCGTCTCAA 1380

AGTACTCCCA GGCGGTGCCA GCCGTGACCG AGGGGCCCAT TCCCGAGGTG CTCAAGAACT 1440

ACATGGACGC CCAGTACTAC GGGGAGATTG GCATCGGGAC GCCCCCCCAG TGCTTCACAG 1500

TCGTCTTCGA CACGGGCTCC TCCAACCTGT GGGTCCCCTC CATCCACTGC AAACTGCTGG 1560

ACATCGCTTG CTGGATCCAC CACAAGTACA ACAGCGACAA GTCCAGCACC TACGTGAAGA 1620

ATGGTACCTC GTTTGACATC CACTATGGCT CGGGCAGCCT CTCCGGGTAC CTGAGCCAGG 1680

ACACTGTGTC GGTGCCCTGC CAGTCAGCGT CGTCAGCCTC TGCCCTGGGC GGTGTCAAAG 1740

TGGAGAGGCA GGTCTTTGGG GAGGCCACCA AGCAGCCAGG CATCACCTTC ATCGCAGCCA 1800

AGTTCGATGG CATCCTGGGC ATGGCCTACC CCCGCATCTC CGTCAACAAC GTGCTGCCCG 1860

TCTTCGACAA CCTGATGCAG CAGAAGCTGG TGGACCAGAA CATCTTCTCC TTCTACCTGA 1920

GCAGGGACCC AGATGCGCAG CCTGGGGGTG AGCTGATGCT GGGTGGCACA GACTCCAAGT 1980

ATTACAAGGG TTCTCTGTCC TACCTGAATG TCACCCGCAA GGCCTACTGG CAGGTCCACC 2040

TGGACCAGGT GGAGGTGGCC AGCGGGCTGA CCCTGTGCAA GGAGGGCTGT GAGGCCATTG 2100

TGGACACAGG CACTTCCCTC ATGGTGGGCC CGGTGGATGA GGTGCGCGAG CTGCAGAAGG 2160

CCATCGGGGC CGTGCCGCTG ATTCAGGGCG AGTACATGAT CCCCTGTGAG AAGGTGTCCA 2220

CCCTGCCCGC GATCACACTG AAGCTGGGAG GCAAAGGCTA CAAGCTGTCC CCAGAGGACT 2280

ACACGCTCAA GGTGTCGCAG GCCGGGAAGA CCCTCTGCCT GAGCGGCTTC ATGGGCATGG 2340

ACATCCCGCC ACCCAGCGGG CCACTCTGGA TCCTGGGCGA CGTCTTCATC GGCCGCTACT 2400

ACACTGTGTT TGACCGTGAC AACAACAGGG TGGGCTTCGC CGAGGCTGCC CGCCTCTAGC 2460

AGCTG 2465

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1448 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapien
(F) TISSUE TYPE: Epithelial

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..6
(D) OTHER INFORMATION: /note= "Restriction site"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1275..1280
(D) OTHER INFORMATION: /note= "Restriction site"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1444..1448
(D) OTHER INFORMATION: /note= "Restriction site"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GAATTCAAGC CCAGAGCCCT GCCATTTCTG TGGGCTCAGG TCCCTACTGC TCAGCCCCTT 60

CCTCCCTCGG CAAGGCCACA ATGAACCGGG GAGTCCCTTT TAGGCACTTG CTTCTGGTGC 120

TGCAACTGGC GCTCCTCCCA GCAGCCACTC AGGGAAAGAA AGTGGTGCTG GGCAAAAAAG 180

GGGATACAGT GGAACTGACC TGTACAGCTT CCCAGAAGAA GAGCATACAA TTCCACTGGA 240

AAAACTCCAA CCAGATAAAG ATTCTGGGAA ATCAGGGCTC CTTCTTAACT AAAGGTCCAT 300

CCAAGCTGAA TGATCGCGCT GACTCAAGAA GAAGCCTTTG GGACCAAGGA AACTTCCCCC 360

TGATCATCAA GAATCTTAAG ATAGAAGACT CAGATACTTA CATCTGTGAA GTGGAGGACC 420

AGAAGGAGGA GGTGCAATTG CTAGTGTTCG GATTGACTGC CAACTCTGAC ACCCACCTGC 480

TTCAGGGGCA GAGCCTGACC CTGACCTTGG AGAGCCCCCC TGGTAGTAGC CCCTCAGTGC 540

AATGTAGGAG TCCAAGGGGT AAAAACATAC AGGGGGGGAA GACCCTCTCC GTGTCTCAGC 600

TGGAGCTCCA GGATAGTGGC ACCTGGACAT GCACTGTCTT GCAGAACCAG AAGAAGGTGG 660

AGTTCAAAAT AGACATCGTG GTGCTAGCTT TCCAGAAGGC CTCCAGCATA GTCTATAAGA 720

AAGAGGGGGA ACAGGTGGAG TTCTCCTTCC CACTCGCCTT TACAGTTGAA AAGCTGACGG 780

GCAGTGGCGA GCTGTGGTGG CAGGCGGAGA GGGCTTCCTC CTCCAAGTCT TGGATCACCT 840

TTGACCTGAA GAACAAGGAA GTGTCTGTAA AACGGGTTAC CCAGGACCCT AAGCTCCAGA 900

TGGGCAAGAA GCTCCCGCTC CACCTCACCC TGCCCCAGGC CTTGCCTCAG TATGCTGGCT 960

CTGGAAACCT CACCCTGGCC CTTGAAGCGA AAACAGGAAA GTTGCATCAG GAAGTGAACC 1020

TGGTGGTGAT GAGAGCCACT CAGCTCCAGA AAAATTTGAC CTGTGAGGTG TGGGGACCCA 1080

CCTCCCCTAA GCTGATGCTG AGCTTGAAAC TGGAGAACAA GGAGGCAAAG GTCTCGAAGC 1140

GGGAGAAGGC GGTGTGGGTG CTGAACCCTG AGGCGGGGAT GTGGCAGTGT CTGCTGAGTG 1200

ACTCGGGACA GGTCCTGCTG GAATCCAACA TCAAGGTTCT GCCCACATGG TCCACCCCGG 1260

TGCAGCCAAT GGCCTCTAGA CAGCTGGCAA GCGGTCCTGC AGACACAGAG GTGATTGTGG 1320

CCTTGGCTGT ATGTGGCTCC ATCCTCTTCC TCCTCATAGT GCTGCTCCTC ACCGTCCTCT 1380

TCCGGATGCA GGCCCAGCCT CCTGGCTACC GCCACGTCGC AGATGGGGAG GACCACGCCT 1440

GAGTCGAC 1448

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1421 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapien
(F) TISSUE TYPE: Epithelial

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..6
(D) OTHER INFORMATION: /note= "Restriction site"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1275..1280
(D) OTHER INFORMATION: /note= "Restriction site"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1416..1421
(D) OTHER INFORMATION: /note= "Restriction site"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GAATTCAAGC CCAGAGCCCT GCCATTTCTG TGGGCTCAGG TCCCTACTGC TCAGCCCCTT 60

CCTCCCTCGG CAAGGCCACA ATGAACCGGG GAGTCCCTTT TAGGCACTTG CTTCTGGTGC 120

TGCAACTGGC GCTCCTCCCA GCAGCCACTC AGGGAAAGAA AGTGGTGCTG GGCAAAAAAG 180

GGGATACAGT GGAACTGACC TGTACAGCTT CCCAGAAGAA GAGCATACAA TTCCACTGGA 240

AAAACTCCAA CCAGATAAAG ATTCTGGGAA ATCAGGGCTC CTTCTTAACT AAAGGTCCAT 300

CCAAGCTGAA TGATCGCGCT GACTCAAGAA GAAGCCTTTG GGACCAAGGA AACTTCCCCC 360

TGATCATCAA GAATCTTAAG ATAGAAGACT CAGATACTTA CATCTGTGAA GTGGAGGACC 420

AGAAGGAGGA GGTGCAATTG CTAGTGTTCG GATTGACTGC CAACTCTGAC ACCCACCTGC 480

TTCAGGGGCA GAGCCTGACC CTGACCTTGG AGAGCCCCCC TGGTAGTAGC CCCTCAGTGC 540

AATGTAGGAG TCCAAGGGGT AAAAACATAC AGGGGGGGAA GACCCTCTCC GTGTCTCAGC 600

TGGAGCTCCA GGATAGTGGC ACCTGGACAT GCACTGTCTT GCAGAACCAG AAGAAGGTGG 660

AGTTCAAAAT AGACATCGTG GTGCTAGCTT TCCAGAAGGC CTCCAGCATA GTCTATAAGA 720

AAGAGGGGGA ACAGGTGGAG TTCTCCTTCC CACTCGCCTT TACAGTTGAA AAGCTGACGG 780

GCAGTGGCGA GCTGTGGTGG CAGGCGGAGA GGGCTTCCTC CTCCAAGTCT TGGATCACCT 840

TTGACCTGAA GAACAAGGAA GTGTCTGTAA AACGGGTTAC CCAGGACCCT AAGCTCCAGA 900

TGGGCAAGAA GCTCCCGCTC CACCTCACCC TGCCCCAGGC CTTGCCTCAG TATGCTGGCT 960

CTGGAAACCT CACCCTGGCC CTTGAAGCGA AAACAGGAAA GTTGCATCAG GAAGTGAACC 1020

TGGTGGTGAT GAGAGCCACT CAGCTCCAGA AAAATTTGAC CTGTGAGGTG TGGGGACCCA 1080

CCTCCCCTAA GCTGATGCTG AGCTTGAAAC TGGAGAACAA GGAGGCAAAG GTCTCGAAGC 1140

GGGAGAAGGC GGTGTGGGTG CTGAACCCTG AGGCGGGGAT GTGGCAGTGT CTGCTGAGTG 1200

ACTCGGGACA GGTCCTGCTG GAATCCAACA TCAAGGTTCT GCCCACATGG TCCACCCCGG 1260

TGCAGCCAAT GGCCTCTAGA CTGCTGGACG AGAACAGCAC GCTGATCCCC ATCGCTGTGG 1320

GTGGTGCCCT GGCGGGGCTG GTCCTCATCG TCCTCATCGC CTACCTCGTC GGCAGGAAGA 1380

GGAGTCACGC AGGCTACCAG ACTATCTAGC CTGGTGTCGA C 1421

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1415 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapien
(F) TISSUE TYPE: epithelial

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..6
(D) OTHER INFORMATION: /note= "Restriction site"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1275..1280
(D) OTHER INFORMATION: /note= "Restriction site"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1410..1415
(D) OTHER INFORMATION: /note= "Restriction site"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GAATTCAAGC CCAGAGCCCT GCCATTTCTG TGGGCTCAGG TCCCTACTGC TCAGCCCCTT 60

CCTCCCTCGG CAAGGCCACA ATGAACCGGG GAGTCCCTTT TAGGCACTTG CTTCTGGTGC 120

TGCAACTGGC GCTCCTCCCA GCAGCCACTC AGGGAAAGAA AGTGGTGCTG GGCAAAAAAG 180

GGGATACAGT GGAACTGACC TGTACAGCTT CCCAGAAGAA GAGCATACAA TTCCACTGGA 240

AAAACTCCAA CCAGATAAAG ATTCTGGGAA ATCAGGGCTC CTTCTTAACT AAAGGTCCAT 300

CCAAGCTGAA TGATCGCGCT GACTCAAGAA GAAGCCTTTG GGACCAAGGA AACTTCCCCC 360

TGATCATCAA GAATCTTAAG ATAGAAGACT CAGATACTTA CATCTGTGAA GTGGAGGACC 420

AGAAGGAGGA GGTGCAATTG CTAGTGTTCG GATTGACTGC CAACTCTGAC ACCCACCTGC 480

TTCAGGGGCA GAGCCTGACC CTGACCTTGG AGAGCCCCCC TGGTAGTAGC CCCTCAGTGC 540

AATGTAGGAG TCCAAGGGGT AAAAACATAC AGGGGGGGAA GACCCTCTCC GTGTCTCAGC 600

TGGAGCTCCA GGATAGTGGC ACCTGGACAT GCACTGTCTT GCAGAACCAG AAGAAGGTGG 660

AGTTCAAAAT AGACATCGTG GTGCTAGCTT TCCAGAAGGC CTCCAGCATA GTCTATAAGA 720

AAGAGGGGGA ACAGGTGGAG TTCTCCTTCC CACTCGCCTT TACAGTTGAA AAGCTGACGG 780

GCAGTGGCGA GCTGTGGTGG CAGGCGGAGA GGGCTTCCTC CTCCAAGTCT TGGATCACCT 840

TTGACCTGAA GAACAAGGAA GTGTCTGTAA AACGGGTTAC CCAGGACCCT AAGCTCCAGA 900

TGGGCAAGAA GCTCCCGCTC CACCTCACCC TGCCCCAGGC CTTGCCTCAG TATGCTGGCT 960

CTGGAAACCT CACCCTGGCC CTTGAAGCGA AAACAGGAAA GTTGCATCAG GAAGTGAACC 1020

TGGTGGTGAT GAGAGCCACT CAGCTCCAGA AAAATTTGAC CTGTGAGGTG TGGGGACCCA 1080

CCTCCCCTAA GCTGATGCTG AGCTTGAAAC TGGAGAACAA GGAGGCAAAG GTCTCGAAGC 1140

GGGAGAAGGC GGTGTGGGTG CTGAACCCTG AGGCGGGGAT GTGGCAGTGT CTGCTGAGTG 1200

ACTCGGGACA GGTCCTGCTG GAATCCAACA TCAAGGTTCT GCCCACATGG TCCACCCCGG 1260

TGCAGCCAAT GGCCTCTAGA AGTGCAGATG ACGACAACTT CCTTGTGCCC ATAGCGGTGG 1320

GAGCTGCCTT GGCAGGAGTA CTTATTCTAG TGTTGCTGGC TTATTTTATT GGTCTCAAGC 1380

ACCATCATGC TGGATATGAG CAATTTTAGG TCGAC 1415

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 1..6
(D) OTHER INFORMATION: /note= "ER retention signal"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Ser Glu Lys Asp Glu Leu
1 5

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..23
(D) OTHER INFORMATION: /note= "PCR Primer P-1 used in construction of sCD4-PCaD"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GAATTCAAGC CCAGAGCCCT GCC

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..23
(D) OTHER INFORMATION: /note= "PCR Primer P-2 used in construction of sCD4-PCaD"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

TCTAGAGGCC ATTGGCTGCA CCG

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..23

(D) OTHER INFORMATION: /note= "PCR Primer P-3 used in construction of sCD4-PCaD"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

TCTAGACTCG TCAGGATCCC GCTG

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..22
(D) OTHER INFORMATION: /note= "PCR Primer P-4 used in construction of sCD4-PCaD"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GTCGACCTAG AGGCGGGCAG CC

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..25
(D) OTHER INFORMATION: /note= "PCR Primer P-1 used in construction of sCD4-HAP"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

TCTAGACAGC TGGCAAGCGG TCCTG

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..24
(D) OTHER INFORMATION: /note= "PCR Primer P-2 used in construction of sCD4-HAP"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GTCGACTCAG GCGTGGTCCT CCCC

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..26
(D) OTHER INFORMATION: /note= "PCR Primer P-1 used in construction of sCD4-L1"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

TCTAGACTGC TGGACGAGAA CAGCAC

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..28
(D) OTHER INFORMATION: /note= "PCR Primer P-2 used in construction of sCD4-L1".

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GTCGACACCA GGCTAGATAG TCTGGTAG

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..27
(D) OTHER INFORMATION: /note= "PCR Primer P-1 used in construction of sCD4-L2"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

TCTAGAAGTG CAGATGACGA CAACTTC

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..30
(D) OTHER INFORMATION: /note= "PCR Primer P-2 used in construction of sCD4-L2"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GTCGACCTAA AATTGCTCAT ATCCAGCATG

We claim:

1. A fusion protein comprising:

a protein which binds to a retroviral envelope protein and

a protein or domain of the protein which targets the fusion protein to a lysosome by causing the protein to be delivered to a lysosome when expressed within a mammalian cell.

2. The protein of claim 1 wherein the binding protein is soluble CD4.

3. The protein of claim 1 wherein the binding protein is selected from the group consisting of sCD4 domain D1 and sCD4 combined domains D1-D2.

4. The protein of claim 1 wherein the targeting protein is selected from the group consisting of procathepsin D (PCaD), human lysosomal membrane protein lamp-1, human lysosomal membrane protein lamp-2, acid phosphatase, and portions thereof targeting the protein to a lysosome.

5. The protein of claim 1 wherein the targeting protein is a lysosomal proenzyme which contains a structural marker for lysosomal targeting.

6. The protein of claim 5 selected from the group consisting of pro-cathepsin D, α-N-Acetylgalactosaminidase, Glycosylasparaginase, Glucocerebrosidase, Procathepsin L, Procathepsin B, and Procathepsin E.

7. The protein of claim 1 wherein the targeting protein is taken from part of the lysosomal membrane proteins, lamp-1 (L1), lamp-2 (L2), and lysosomal acid phosphatase (HAP).

8. The protein of claim 1 wherein there are multiple targeting proteins.

9. The protein of claim 8 wherein the fusion protein is selected from the group consisting of sCD4-PCaD-HAP and sCD4-PCaD-PCaD.

10. The protein of claim 1 wherein the targeting protein and the binding protein are separated by a peptide sequence.

11. A gene encoding a fusion protein comprising:

a protein which binds to a retroviral envelope protein and

a protein or domain of the protein which targets the fusion protein to a lysosome by causing the protein to be delivered to a lysosome when expressed within a mammalian cell.

12. The gene of claim 11 encoding a protein selected from the group consisting of sCD4-PCaD, sCD4-HAP, sCD4-L1, and sCD4-L2.

13. The gene of claim 11 encoding a fusion protein wherein the binding protein is soluble CD4.

14. The gene of claim 11 encoding a fusion protein wherein the binding protein is selected from the group consisting of sCD4 domain D1 and sCD4 combined domains D1-D2.

15. The gene of claim 11 encoding a fusion protein wherein the targeting protein is selected from the group consisting of procathepsin D (PCaD), human lysosomal membrane protein lamp-1, human lysosomal membrane protein lamp-2, acid phosphatase, and portions thereof targeting the protein to a lysosome.

16. The gene of claim 11 encoding a fusion protein wherein the targeting protein is a lysosomal proenzyme which contains a structural marker for lysosomal targeting.

17. The gene of claim 11 encoding a fusion protein selected from the group consisting of pro-cathepsin D, α-N-Acetylgalactosaminidase, Glycosylasparaginase, Glucocerebrosidase, Procathepsin L, Procathepsin B, and Procathepsin E.

18. The gene of claim 10 encoding a fusion protein wherein the targeting protein is taken from part of the lysosomal membrane proteins, lamp-1 (L1), lamp-2 (L2), and lysosomal acid phosphatase (HAP).

19. The gene of claim 10 encoding a fusion protein wherein there are multiple targeting proteins.

20. The gene of claim 10 encoding a fusion protein wherein the fusion protein is selected from the group consisting of sCD4-PCaD-HAP and sCD4-PCaD-PCaD.

21. The gene of claim 11 further comprising a vector.

22. A method for treating a viral disease comprising introducing into cells that are infected or exposed to a retrovirus a gene encoding a fusion protein comprising:

a protein which binds to a retroviral envelope protein and

a protein or domain of the protein which targets the fusion protein to a lysosome by causing the protein to be delivered to a lysosome when expressed within a mammalian cell.

23. The method of claim 22 wherein the virus is human immunodeficiency virus.

24. The method of claim 22 wherein the gene is introduced into the cells within a viral vector.

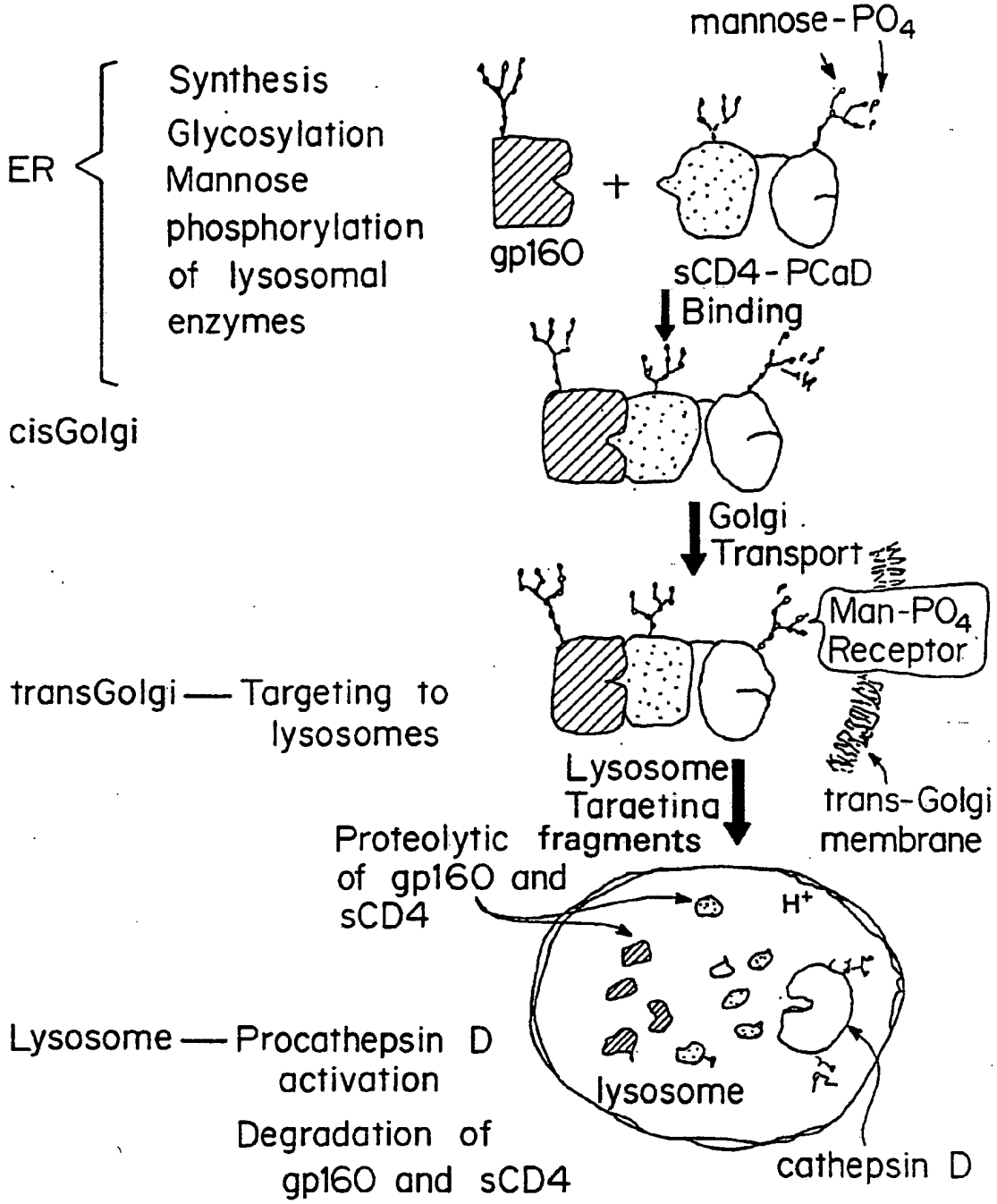
25. The method of claim 22 wherein the gene encodes a protein selected from the group consisting of sCD4-PCaD, sCD4-HAP, sCD4-L1, and sCD4-L2.

26. The method of claim 22 wherein the gene encodes a fusion protein wherein the binding protein is selected from the group consisting of sCD4, sCD4 domain D1 and sCD4 combined domains D1-D2.

27. The method of claim 22 wherein the gene encodes a fusion protein wherein the targeting protein is selected from the group consisting of procathepsin D (PCaD), human lysosomal membrane protein lamp-1, human lysosomal membrane protein lamp-2, acid phosphatase, pro-cathepsin D, $\alpha$-N-Acetylgalactosaminidase, Glycosylasparaginase, Glucocerebrosidase, Procathepsin L, Procathepsin B, and Procathepsin E, and portions thereof targeting the protein to a lysosome.

28. The method of claim 22 wherein the gene encodes a fusion protein including multiple targeting proteins.

FIG. 1


mannose-PO4
ER
Synthesis
Glycosylation
Mannose
phosphorylation
of lysosomal
enzymes
cisGolgi
gp160
+
sCD4-PCaD
Binding
Golgi
Transport
Man-PO4
Receptor
transGolgi — Targeting to
lysosomes
Lysosome
Targeting
trans-Golgi
membrane
Proteolytic fragments
of gp160 and
sCD4
H+
Lysosome — Procathepsin D
activation
Degradation of
gp160 and sCD4
lysosome
cathepsin D

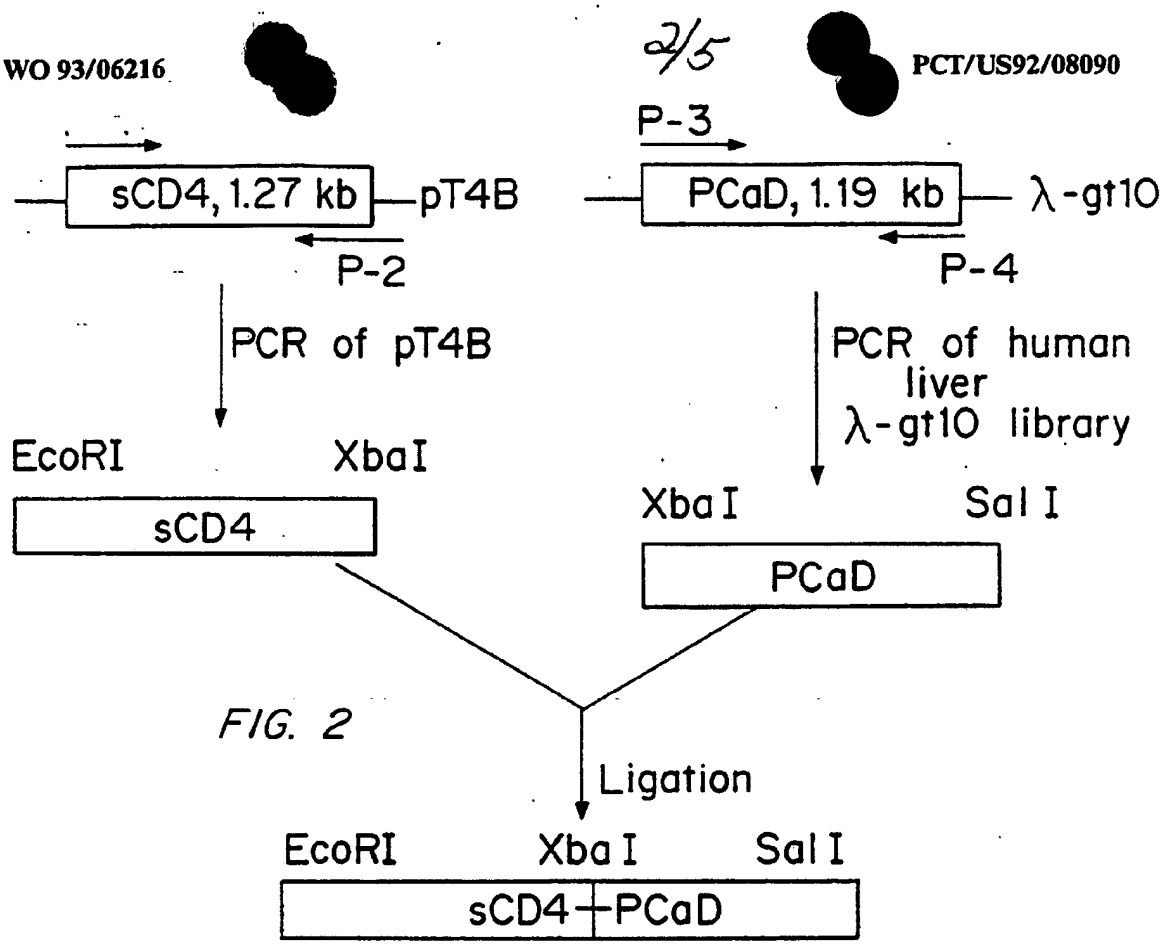
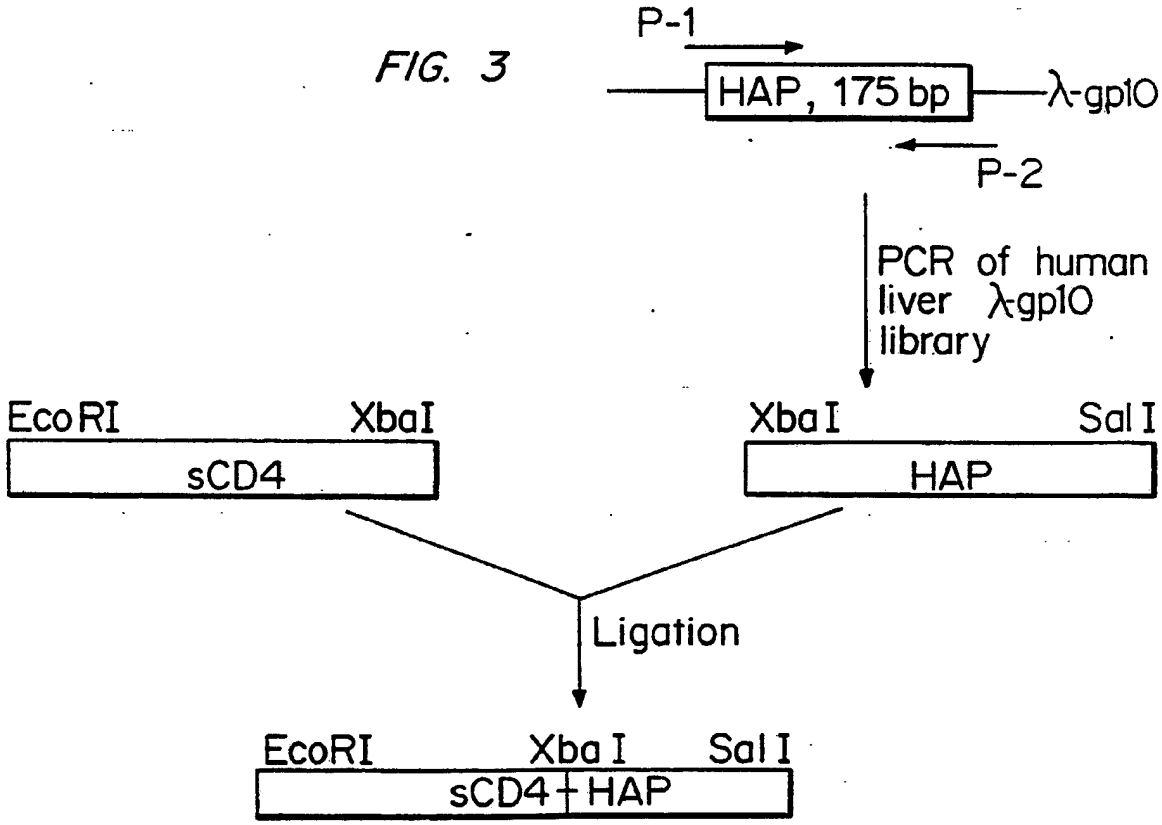

P-3
sCD4, 1.27 kb
pT4B
PCaD, 1.19 kb
λ-gt10
P-2
P-4
PCR of pT4B
PCR of human liver λ-gt10 library
EcoRI
XbaI
sCD4
XbaI
SalI
PCaD
Ligation
EcoRI
XbaI
SalI
sCD4+PCaD

FIG. 2

FIG. 3


P-1
HAP, 175 bp
λ-gp10
P-2
PCR of human liver λ-gp10 library
EcoRI
XbaI
sCD4
XbaI
SalI
HAP
Ligation
EcoRI
XbaI
SalI
sCD4+HAP

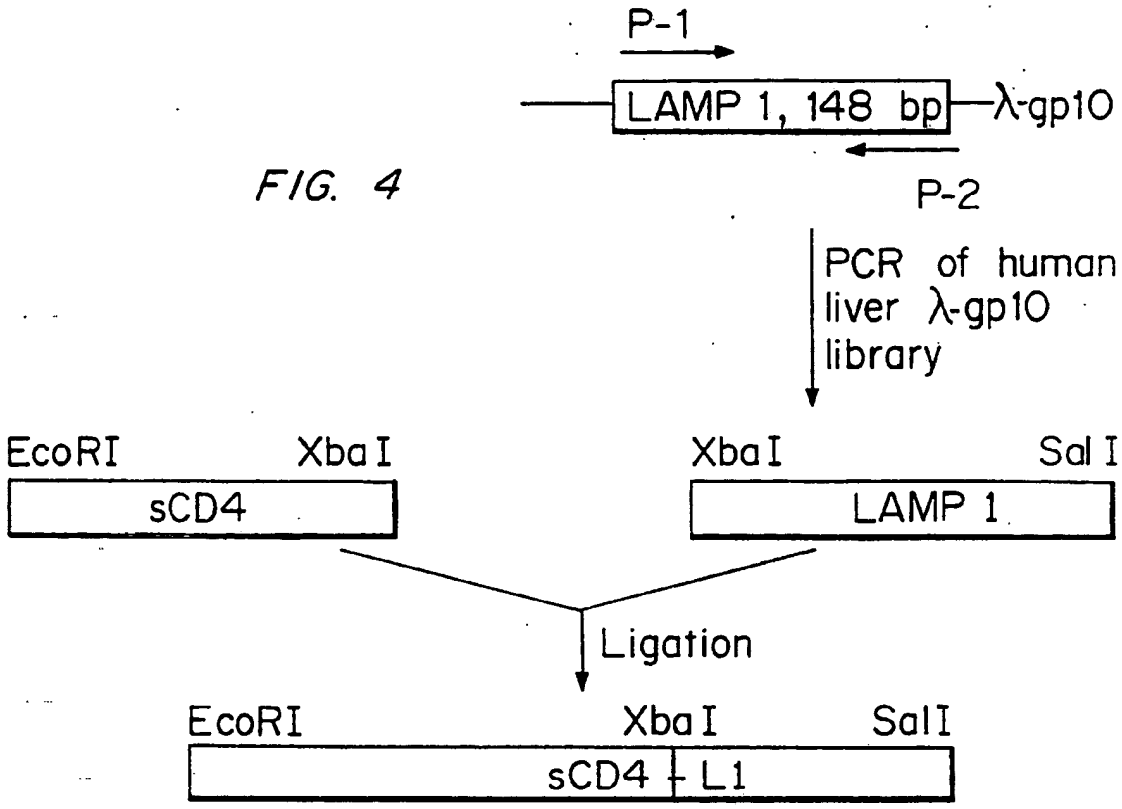

P-1
LAMP 1, 148 bp
λ-gp10
P-2
PCR of human liver λ-gp10 library
EcoRI
Xba I
sCD4
Xba I
Sal I
LAMP 1
Ligation
EcoRI
Xba I
Sal I
sCD4
L1

FIG. 4

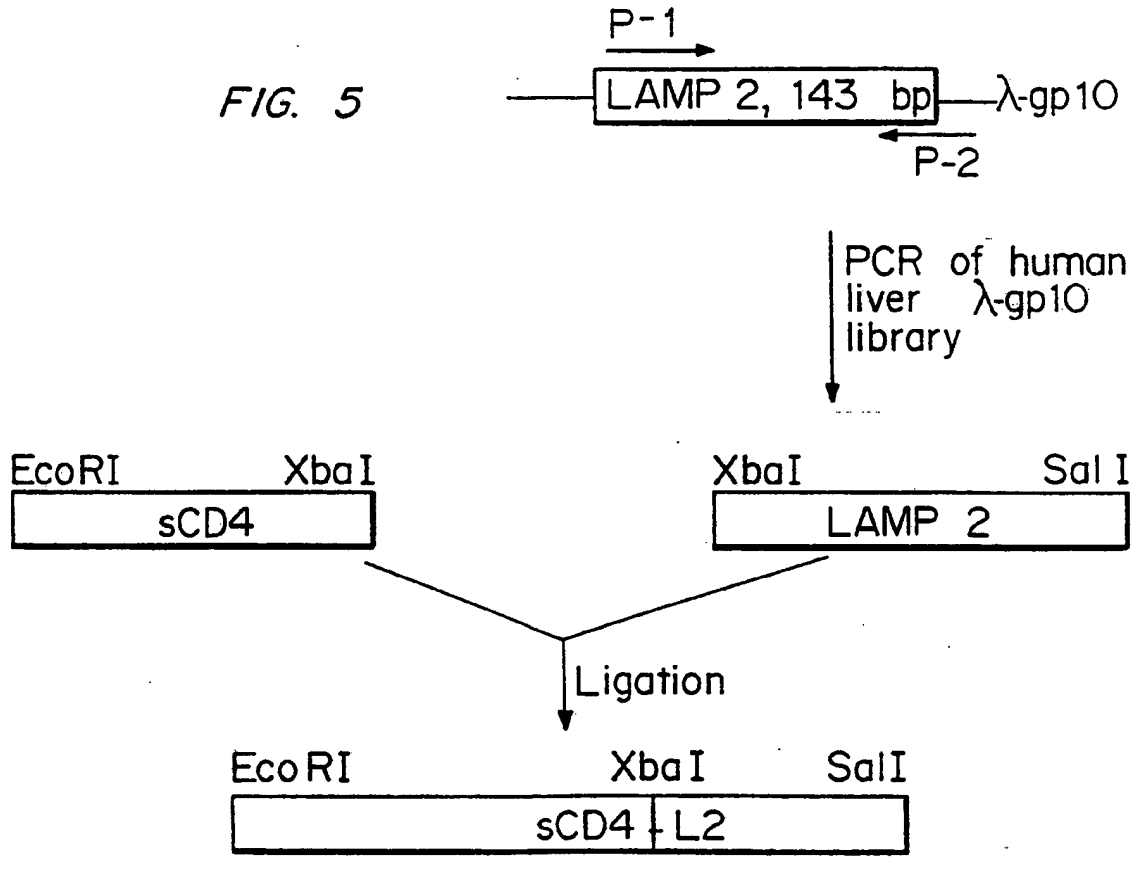

P-1
LAMP 2, 143 bp
λ-gp10
P-2
PCR of human liver λ-gp10 library
EcoRI
Xba I
sCD4
Xba I
Sal I
LAMP 2
Ligation
EcoRI
Xba I
Sal I
sCD4
L2

FIG. 5

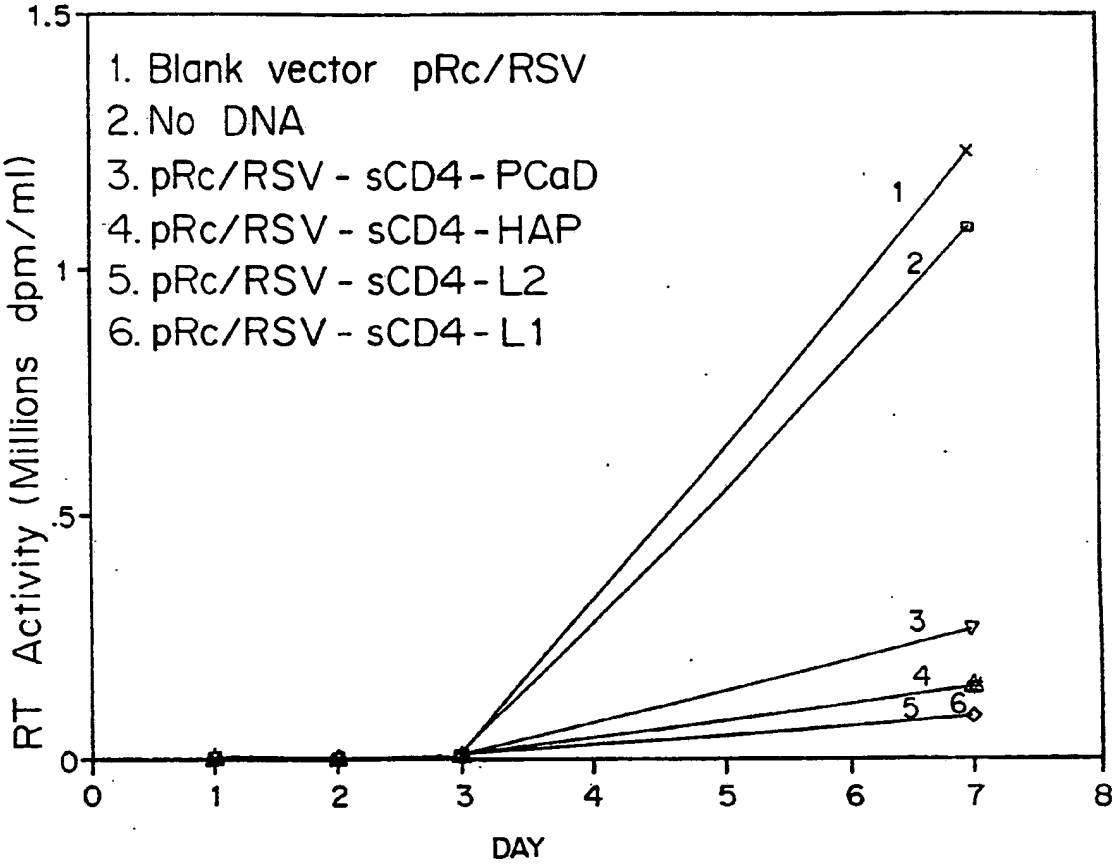

1. Blank vector pRc/RSV
2. No DNA
3. pRc/RSV - sCD4 - PCaD
4. pRc/RSV - sCD4 - HAP
5. pRc/RSV - sCD4 - L2
6. pRc/RSV - sCD4 - L1
RT Activity (Millions dpm/ml)
1.5
1
.5
0
0 1 2 3 4 5 6 7 8
DAY
1
2
3
4
5
6

FIG. 6

FIG. 7a

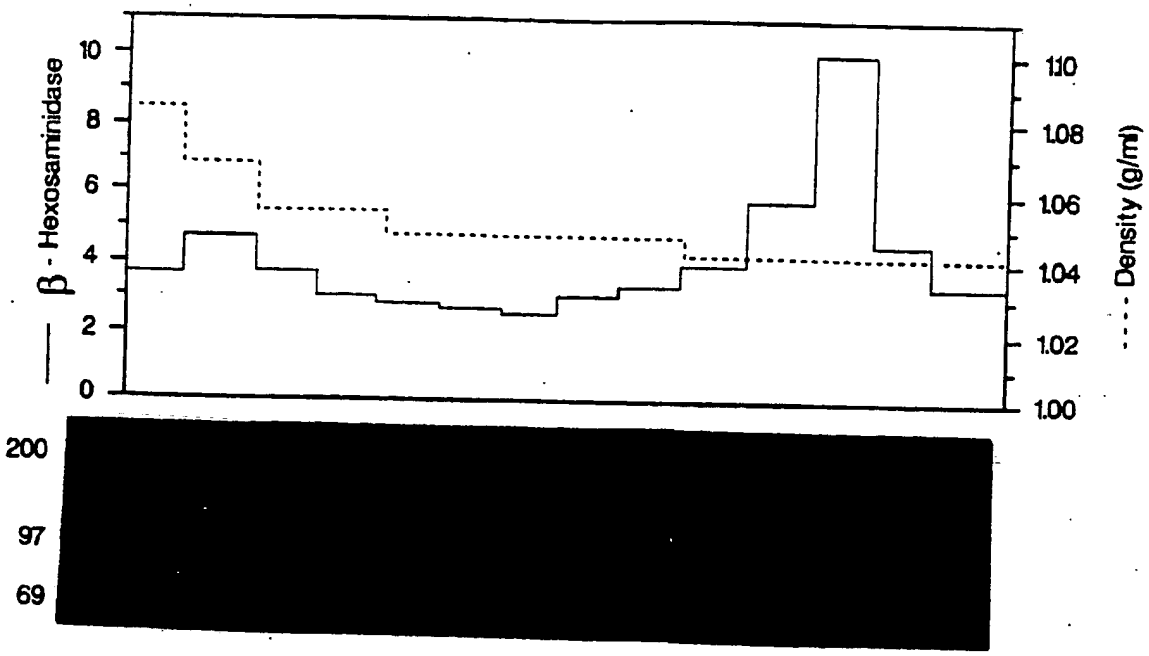
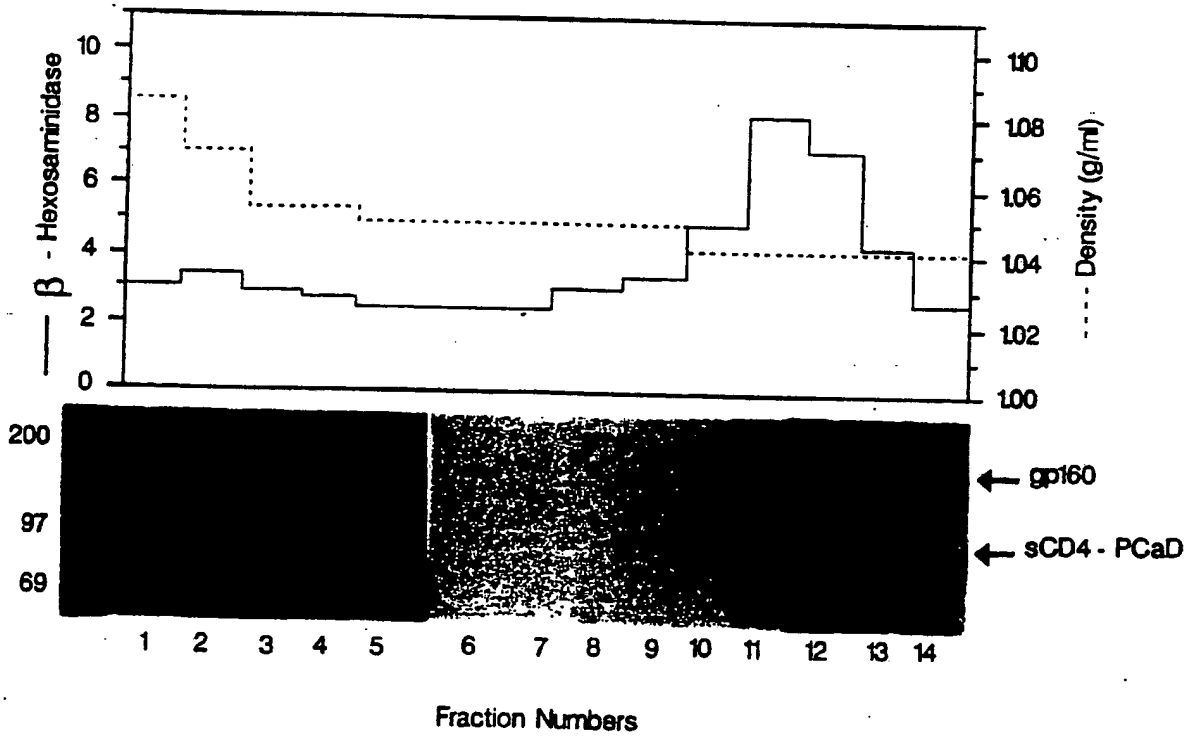



FIG. 7b



# INTERNATIONAL SEARCH REPORT

International Application No PCT/US 92/08090

## I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)[6]

According to International Patent Classification (IPC) or to both National Classification and IPC

Int.Cl. 5 C12N15/12; C12N15/62; C07K13/00; C12N9/16
C12N9/18; C12N9/24; C12N9/64; A61K37/02

## II. FIELDS SEARCHED

Minimum Documentation Searched[7]

| Classification System | Classification Symbols |
|---|---|
| Int.Cl. 5 | C07K ; C12N ; A61K |

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched[8]

## III. DOCUMENTS CONSIDERED TO BE RELEVANT[9]

| Category[10] | Citation of Document,[11] with indication, where appropriate, of the relevant passages[12] | Relevant to Claim No.[13] |
|---|---|---|
| Y | WO,A,9 004 414 (BIOGEN INC., US)<br>3 May 1990<br><br>See the claims | 1-7,<br>10-18,<br>21-27 |
| Y | EMBO JOURNAL.<br>vol. 9, no. 11, 1990, EYNSHAM, OXFORD GB<br>pages 3497 - 3506<br>Peters, Christoph et al.; 'Targeting of a lysosomal memebrane protein : a tyrosine -containing endocytosis signal in the cytoplasmic tail of lysosomal acid phosphatase is necessary and sufficient for targeting to lysosomes.'<br>cited in the application<br>see the whole document | 1-7,<br>10-18,<br>21-27 |

-/--

[10] Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

## IV. CERTIFICATION

| Date of the Actual Completion of the International Search | Date of Mailing of this International Search Report |
|---|---|
| 15 JANUARY 1993 | 04. 02. 93 |
| **International Searching Authority** | **Signature of Authorized Officer** |
| EUROPEAN PATENT OFFICE | S.A. NAUCHE |

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

| Category ° | Citation of Document, with indication, where appropriate, of the relevant passages | Relevant to Claim No. |
|---|---|---|
| Y | COMPTE RENDU DE L'ACADEMIE DES SCIENCES DE PARIS<br>vol. 308, no. III, 1989, PARIS, FR<br>pages 401 - 406<br>Clément, J.M. et al.; 'Propriétés neutralisantes pour le virus HIV d'une protéine hybride MalE-CD4 exprimée chez E.coli et purifiable en une étape.'<br>see the whole document | 1-7, 10-18,21 |
| Y | BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS.<br>vol. 164, no. 3, 15 November 1989, DULUTH, MINNESOTA US<br>pages 1113 - 1120<br>Nogushi, Y. et al.; 'Isolation and sequencing of a cDNA clone encoding 96 kDa sialoglycoprotein in rat liver lysosomal membranes.'<br>see the whole document | 1-5,7, 10-16, 18,21-27 |
| A | JOURNAL OF BIOLOGICAL CHEMISTRY. (MICROFILMS)<br>vol. 265, no. 35, 15 December 1990, BALTIMORE, MD US<br>pages 21859 - 21886<br>Wang, Anne M. et al.; 'Human alpha-N-acetylgalactosaminase molecular cloning, nucleotide sequence, and expression of a full-length cDNA.' | |

# ANNEX TO THE INTERNATIONAL SEARCH REPORT ON INTERNATIONAL PATENT APPLICATION NO.

US 9208090
SA 65465

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report.
The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information. 15/01/93

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---|---|---|
| WO-A-9004414 | 03-05-90 | CA-A- 2000989 | 18-04-90 |